② LEVEL II

JTR-79-07

MULTI-PLATFORM DETECTION STUDY

PREPARED FOR
SIGNAL PROCESSING BRANCH
NAVAL SURFACE WEAPONS CENTER, WHITE OAK
SILVER SPRING, MARYLAND 20910

FINAL REPORT
N60921-79-C-0120

AUGUST 1979

J.S. LEE ASSOCIATES INC.
2001 Jefferson Davis Highway, Suite 201
Arlington, Virginia 22202

MULTI-PLATFORM DETECTION STUDY

by

Leonard E. Miller

JTR-79-07

August 1979

Prepared for:

Signal Processing Branch, Code U22
Naval Surface Weapons Center
White Oak Laboratory
Silver Spring, Maryland 20910

Contract N60921-79-C-0120

Prepared by:

J. S. LEE ASSOCIATES, INC.
2001 Jefferson Davis Highway
Suite 201, Crystal Plaza 1
Arlington, Virginia 22202
(703) 979-2230

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 Multi-Platform Detection Study.

5. TYPE OF REPORT & PERIOD COVERED

9 Final rept. May-Aug[redacted]79

6. PERFORMING ORG. REPORT NUMBER

14 JTR-79-07

7. AUTHOR(s)

10 Leonard E./Miller

8. CONTRACT OR GRANT NUMBER(s)

15 N60921-79-C-0120

9. PERFORMING ORGANIZATION NAME AND ADDRESS

J. S. LEE ASSOCIATES, INC.
2001 Jefferson Davis Highway
Arlington, Virginia 22202

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

64261N; 16 W0480
W0480AS; 9U31DF.

11. CONTROLLING OFFICE NAME AND ADDRESS

Code U22, Naval Surface Weapons Center
White Oak Laboratory
Silver Spring, Maryland 20910

12. REPORT DATE

11 Aug[redacted]79

13. NUMBER OF PAGES

75 + vi

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

12 82p.

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

Lee (J.S.) Associates, Inc
Arlington, VA

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

17 W0480

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Detection, multiplatform, multisensor, estimation, adaptive, covariance, multivariate analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Classical (separate) and joint detection and estimation procedures are extended to the case of complex multidimensional data. Initial results for complex multivariate Gaussian processes indicate that the detector structure involves determinants of sample covariance matrices--a mixture of conventional power and correlation detectors.

393 892 Gw

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

MULTI-PLATFORM DETECTION STUDY

TABLE OF CONTENTS

Glossary of Notation

A	covariance matrix $(m \times m)$
$B(\cdot,\cdot)$	beta function
C	sample covariance matrix $(m \times m)$
$cof(b_{ij})$	cofactor of matrix element b_{ij}
D/E	"detection and estimation"
$E\{\cdot\}$	mathematical expection
etr (B)	$\exp\{tr\ B\}$
H_0, H_1	hypotheses
I	identity matrix
$K_\nu(b)$	modified Bessel function of the second kind
$K_\nu^{(m)}(B)$	matrix argument $(m \times m)$ Bessel functions
$L(\cdot)$	likelihood ratio
m	number of sensors or data channels
$\underline{m}, M$	mean vector $(m \times 1)$ and matrix expansion $(m \times n)$
ML	"maximum likelihood"
n	number of samples
$p(\cdot)$	pdf
pdf	"probability density function"
q	$=(m+1)/2$
QCF	"quadratic cost function"
R	risk or average cost
ROC	"receiver operating characteristics"
SCF	"simple cost function"
trB	trace of the matrix B
$\underline{u}, U$	in-phase data vector, matrix (real part)
$\underline{v}, V$	quadrature data vector, matrix (imaginary part)

X	(complex) data matrix $(m \times n)$
X_0	matrix expansion of $\underline{\mu}$ $(m \times n)$
x_n	n:th sample
X_n	samples up to and including x_n
$\Gamma(\cdot)$	gamma function
$\Gamma_m(\cdot)$	generalized gamma function (see 4-19)
Γ	data space
Δ	unit delay
θ	set of signal parameters
η	set of noise parameters
Λ	likelihood ratio
λ	threshold
$\underline{\mu}$	sample mean vector $(m \times 1)$
π_0, π_1	a priori probabilities
$\underline{b}'$, B'	transpose of vector $\underline{b}$, matrix B
$\underline{b}^*$, B^*	complex conjugate transpose
b^{ij}	element of B^{-1}
$\|B\|$	determinant of matrix B

List of Figures and Tables

1.0 INTRODUCTION

1.1 Background

The physical problem which forms the backdrop for the work summarized in this report may be described as follows: A definite number (m) of discrete, passive sensors whose locations are known have been deployed in such a way as to enable them to sense emissions due to a source of interest if it happens to be within a certain area. It is desired to process the data obtained from these sensors on a given observation interval in a manner that permits statements to be made with a high degree of confidence concerning the absence or presence of the source and its location as a function of time.

While the technology for performing these tasks jointly is rather mature for the special case in which the sensors are quite near each other (within a wavelength) and of the same type, the procedures to be followed in the general case have rarely been developed to the point of operational capability. For example, arrays of sensors physically connected to one another have long been used to couple source direction with detection. Most often, however, systems designed to deal with multiple sensors assume the posture illustrated in Figure 1-1.

In this conceptual diagram the desired information (localization parameters) is shown as being the result of a regression (model-fitting) involving estimates of parameters directly related to "preprocessing" of the received data; the phrase "with memory" indicates tracking. For example, if the extracted parameters are bearings, then the regression may simply be a position "fix"; use of an extended Kalman filter or other tracking algorithm allows other information, such as velocity, to be estimated by remembering past fixes and calculating trends. The role of detection in this system is to help the operator to select

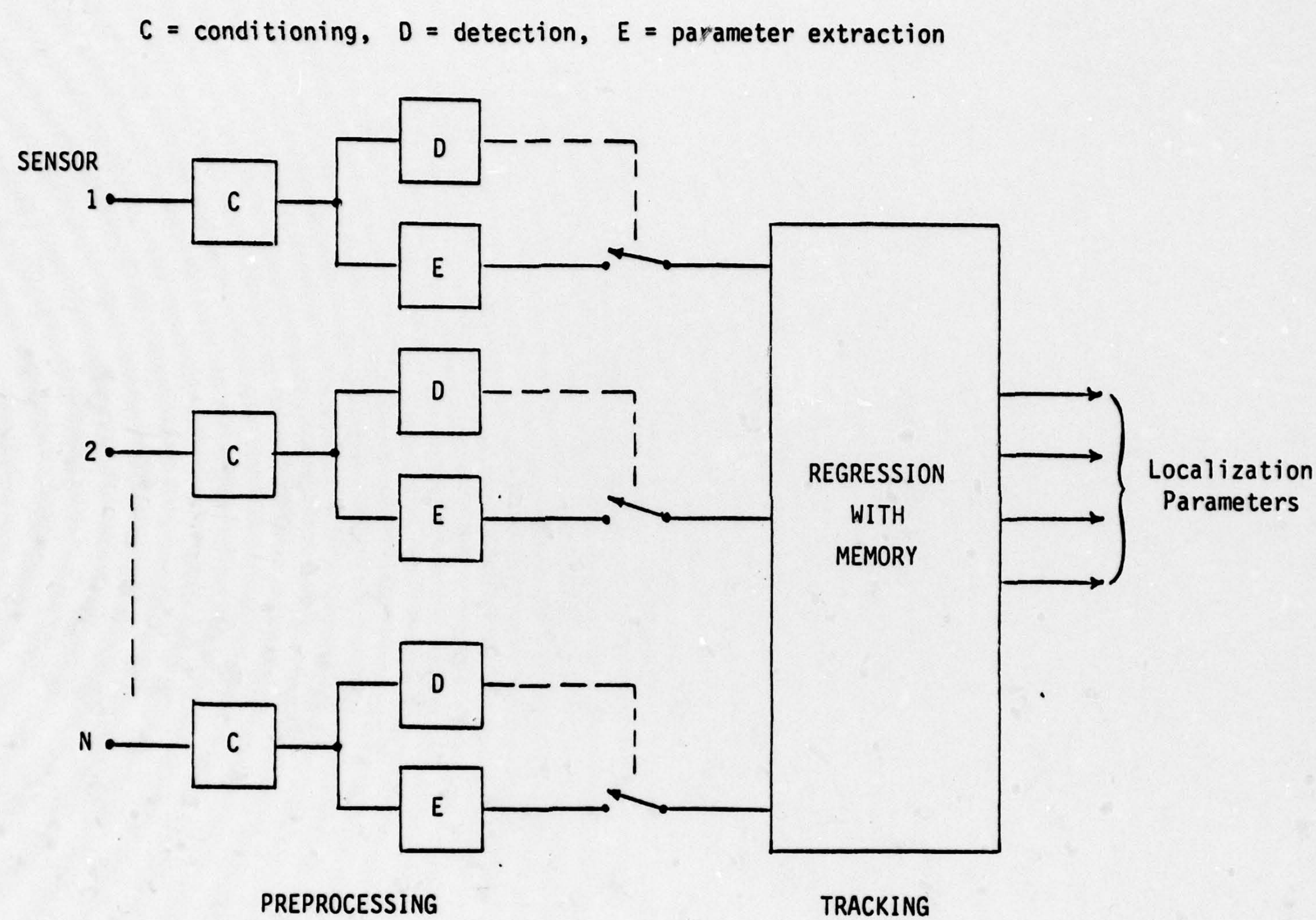


FIGURE 1-1 SYSTEM DIAGRAM

sensors which have "good data" on them in the first place; usually, "energy detection" is implied, although "SNR detection" may occasionally be employed.

The "conditioning" function included in the figure refers to operations such as bandpass filtering, time delays to accomplish "steering", and transmission over a data communications link. The entire diagram may be digital or analog, and conditioning might entail sampling, A/D conversion, and FFT processing as well.

Having set up this figurative system as a reference, several of its (typical) features are worth noting as a means to introduce ideas which are pursued in this report. The development of these ideas or concepts, as applied in the present physical problem, actually began several years ago and is now beginning to produce results. Originally, the question was asked, "How can the detection/estimation outcomes at one sensor (with high SNR) be used to aid or 'coach' those at another sensor (with low SNR)?" This question arose in the context of a system which in its essential aspects is described very well by Figure 1-1, with the preprocessing performed at different physical sites or "platforms."

Modular vs. Multidimensional

One way to describe what is illustrated in the figure is to say, "The preprocessing is modular." That is, detection and signal parameter extraction operations are performed on the data from each sensor separately, in isolation or remote from the data received by the other sensors. This situation existed for the excellent reason that each sensor-bearing platform needed the capability to perform these functions solo. How then to

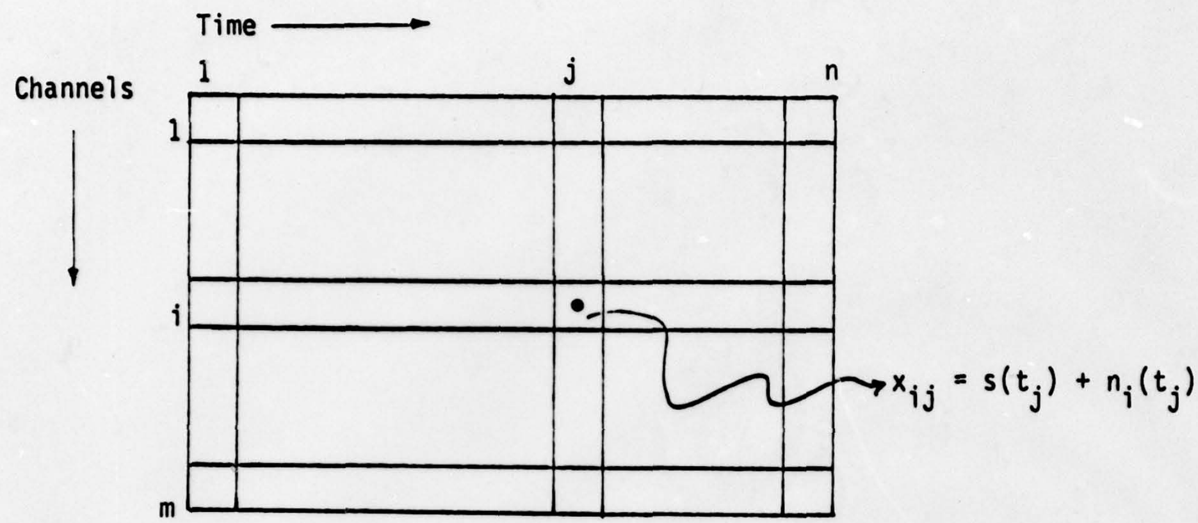
join capabilities when operating in consort, was the question. At first, the commitment to modularity was such a big factor that solutions were sought in which preprocessing results at one platform are used in some way to improve or to help obtain those at another. This can be done a number of ways; however, it is clear that there are better ways to approach the problem.

Consider the "data matrix" of Figure 1-2. Each row consists of time samples from a different sensor. (The samples could be spectral samples - the idea is the same.) What the modular preprocessing in Figure 1-1 does is treat each row of this matrix data base separately. In general, there is information contained in the inter-relations of the samples from row to row. Therefore, we expect that preprocessing that operates on the data as a whole will obtain better results. That is, a multidimensional approach seems to be called for alongside, if not replacing, the modular preprocessing.

Partitioned vs. Coupled

Another description of the system in Figure 1-1 is given by the statement, "The detection and parameter extraction functions are partitioned." Unless the detector is nonparametric, or "distribution-free," in form, then, it makes decisions entirely upon the basis of *a priori* information - the "known parameter" solution --or upon what analysis or experience has indicated the marginal distribtuion of the data should be (again a function of *a priori* information only). Actual detectors hardly ever are built this way since noise and signal power very so widely in physical problems. Instead, detection algorithms designed for composite or variable parameter hypotheses usually include processing designed to estimate one or more key

DATA MATRIX



$\{X_{ij}\}$: mn Data Samples of, say,

m Jointly Gaussian Random Processes.

FIGURE 1-2 DATA MATRIX

parameters, or at least to choose values which satisfy certain criteria such as "worst case." So then, in practice detectors usually include some form of estimation.

The form of estimation that detectors employ may not necessarily be the same as that needed for parameter extraction, however, so that functionally the partitioning shown in Figure 1-1 may still be said to exist. The reasons for this may be explained with the help of Figure 1-3 Conceptually, detection is the process of selecting the most likely hypothesis, the hypotheses being probability distribution models depending upon such parameters as mean and variance (which in turn are related to signal and noise parameters). Parameter extraction may be seen as the process of fitting (or regressing) values of unknown parameters to the data through an assumed dependency or model.

There is an obvious similarity of form between the two processes: both conceptually, at least, involve regressions on the data; the parameters varied in either process are related. The chief difference between them seems to be in the criteria which the regressions must satisfy. However, if both tasks are to be performed, we should like to optimally perform both simultaneously. This goal has motivated research in what the literature calls "coupled" or joint detection and estimation.

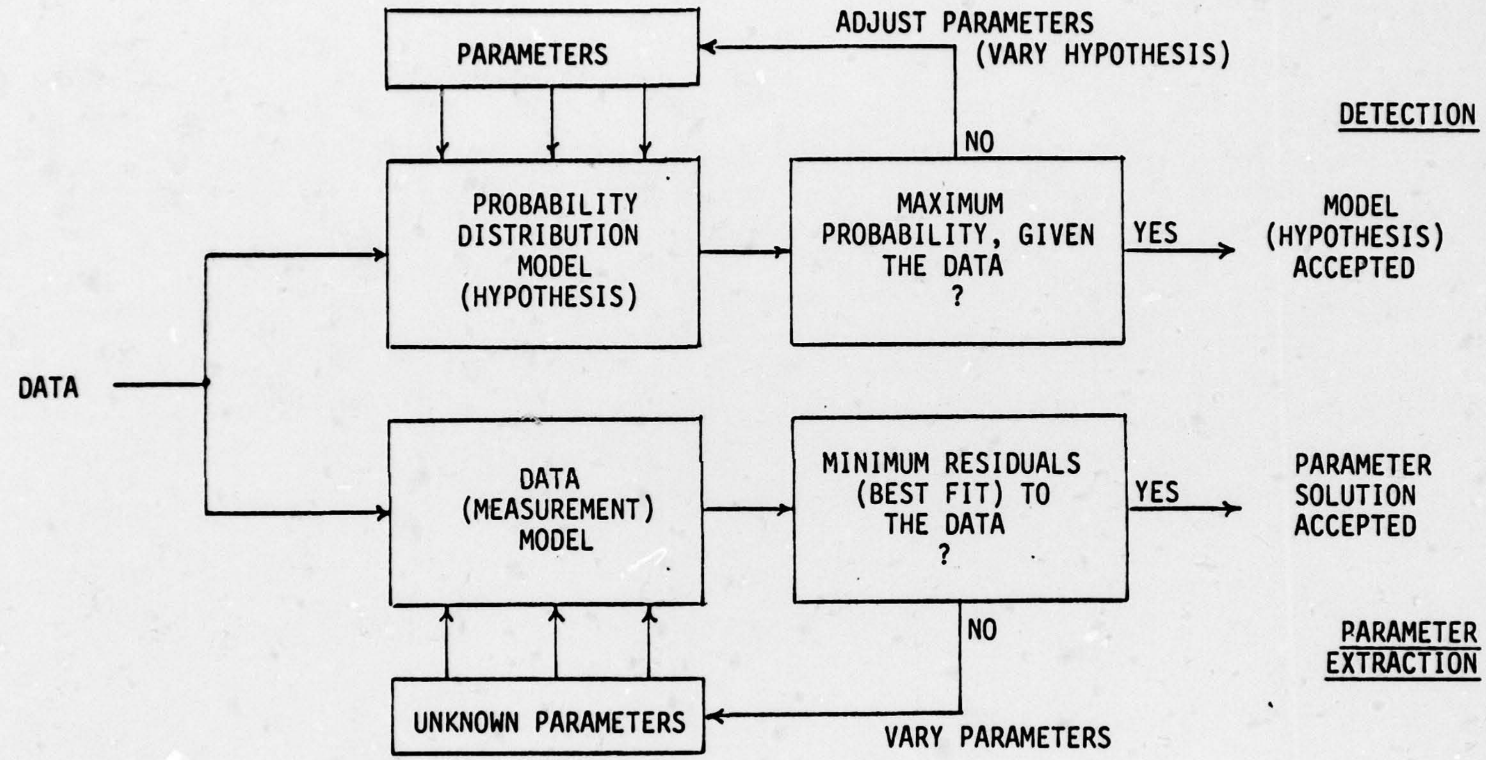


FIGURE 1-3 CONCEPTUAL DETECTION AND PARAMETER EXTRACTION

1.2 Scope of the Study

This report summarizes a four man-month contractual effort to pursue the concepts just discussed. Basically, the work described herein is exploratory--that is to say, it was begun with rather broad and partially defined questions in mind. Therefore, although some results were obtained which are suitable for immediate simulation connected with exploratory system development, the bulk of the study's output is in the form of further questions which are more definite than those which motivated the work.

1.2.1 Literature Review

In its essence, the work has been an effort to discover and to summarize what the technical literature has to offer the engineer who is attempting to deal with the physical problem and concepts that have been described, synthesizing contributions from different disciplines as required. For example, Appendix A is an early effort to apply the techniques of analysis of variance to the multisensor problem (predating the contract). Synthesis is required because typically only a portion of the problem is treated in one place. The short bibliography included with this report is an indication of the variety of sources which have been found to contain useful material.

There is undoubtedly much information in the mathematics and statistics literature which would be helpful to the signal processing engineering community. Extracting this information is difficult because often the engineer lacks the background necessary to ask the right questions or to interpret the significance of what he finds.

Under the limited level of effort which is summarized herein, a beginning has been made.

Literature which seems to have potential for answering the questions we are dealing with can be recognized by the following key words:

multivariate statistical analysis
empirical Bayesian procedures
multi-parameter pattern recognition
functions of multiple arguments
multiple time series analysis

This is not to say that every article dealing with these topics is useful--quite often the "interesting part" is rendered less useful by simplifying assumptions and analytical conventions which do not apply to the signal processing problem. For example, the results we seek are for complex (narrowband) data, whereas the literature mostly treats real data.

Literature which has not yielded guidance for the multi-sensor problem in the sense that it is defined here include that which deals with

multidimensional systems theory
analysis of variance (as applied to the life sciences).

1.2.2 Approach Taken

While the literature review process necessarily will take an extended period as we learn what exists and how to interpret it, along the way it is desirable to test the procedures and analytical tools found which seem promising. Thus in this report efforts to formulate a unified approach to the multi-sensor signal processing--specifically, detection and estimation (D/E)--are summarized as they have developed under the contract.

The structure of the report is the following: Chapter 2 is a review of classical D/E given partly to provide a self-contained context for the later chapters, and partly as a means for introducing a vocabulary and a notation.

Chapter 3 expands on two basic lines of approach to combined D/E, for the case of one channel or sensor (as they appear in the literature). These approaches are, it seems, the closest to dealing with the sort of questions that have been posed that we have found so far.

Chapter 4 then summarizes formulation of the multisensor D/E problem as a "matrix data" problem, and shows how the classical and combined D/E theories look when applied to more than one data channel.

The report "flow" can be diagrammed in the following way:

	single channel	m channels
separate D/E	Chapter 2 (review)	Chapter 4 (extension)
combined D/E	Chapter 3 (adaptation)	

Since only a modest level of effort was involved, a great deal of what seems to be promising work remains to be done. In Chapter 5, the results of this exploratory study are discussed in terms of the interpretations which are provided for the design of multisensor processors. Also, recommendations are given for applications and for further studies.

2.0 MODELING AND ANALYTICAL APPROACHES

In this chapter we establish the framework for the effort summarized in this report. First, the analytical models employed throughout the work are described, including notation. Then the operations of detection and estimation are defined in the usual way, and the classical (single channel) results summarized by way of review. Finally, the procedures to follow in the subsequent chapters are given.

2.1 System Description

As mentioned at the very beginning of this report, presumably there exist at least potentially in the medium being considered (e.g. underwater) waveforms due to the source or sources of interest. We restrict our attention to a single source, whose waveform we denote by $s(t;\theta)$ to indicate variation in time and dependence upon certain parameters θ. Whether this source of interest is present or not, the medium is such that there exists at each of m sensors a noise waveform $n_i(t,\eta_i)$, $i=1,2...,m$; the noise parameters $\{\eta_i\}$ are in general different in value at each sensor.

By "sensor" we shall refer in this work to whatever appropriate transducers and conditioning may be required to acquire data, including a certain amount of processing whose nature will be specified in particular examples as they are brought up. The physical locations of the sensors are taken to be different, so that observation in both space and time is performed by the collection of sensors, whose outputs are assumed to be available to a centralized processor. Our primary concern is with the structure of this processor; therefore, our modelling

effort begins with the "data" $\{x_i(t)\}$ from the m channels, as illustrated in Figure 2-1.

In effect, we are placing the "observer" inside the box marked "centralized processor." (This, of course, is the standard operating procedure for analysis of this type). Having been placed in these circumstances, the observer is going to try to perform the assigned system tasks as well as possible with the data supplied. Basically the tasks (described in detail later) are to make inferences based on the numerical data concerning the source of interest. Depending upon the task, certain information needs to be supplied to, or developed by, the observer. Specifying this information amounts to modelling the data.

For example, in order to decide whether the source of interest ("signal") is present, we should like to know the *probability* distributions of the data under the hypotheses

$$H_0 : \text{noise only} \qquad H_1 : \text{signal and noise.} \tag{2-1}$$

The appropriate probability density functions (pdf's) are written

$$p_i(X) \equiv p(X|H_i), \; i = 0, 1 \tag{2-2}$$

and where the functional forms of pdf's are in general different for arguments and conditions, according to the notation used in this report. Here also we use the notation X to refer to all the data observed on a given interval. It is assumed that, with respect to a given bandwidth, the data are in the form of complex numbers (in-phase and quadrature components) and may be taken together to form a data matrix:

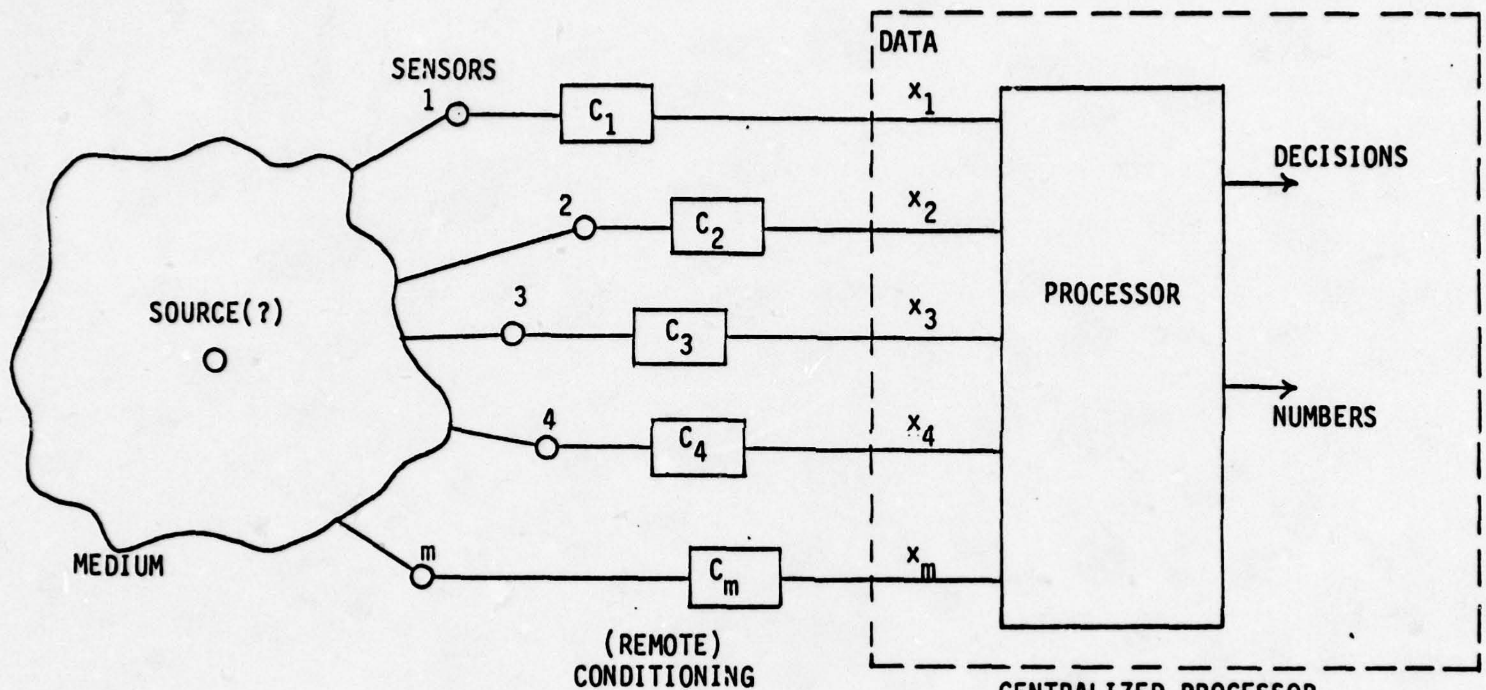


FIGURE 2-1 MULTISENSOR SYSTEM MODEL

$$X = \{x_{ik}\}; \quad i = 1, 2, \ldots m; \quad k=1, 2, \ldots n; \tag{2-3}$$

where

$$x_{ik} = x_i(t_k) = u_{ik} + jv_{ik}. \tag{2-4}$$

The "marginal" or unconditioned pdf's $p_0(X)$ and $p_1(X)$ are not usually available; instead, one has pdf's "indexed" or conditioned by signal and noise parameters:

$$\begin{aligned} p_1(X|\theta,\eta) &\equiv p(X|\theta,\eta) \\ p_0(X|\eta) &\equiv p(X|\eta). \end{aligned} \tag{2-5}$$

If the values of the parameters θ and η are known, then the pdf's of (2-5) are simply "parameterized" versions of those of (2-2). If not, then the values either need to estimated in some way or the <u>a priori</u> pdf $p(\theta,\eta)$ specified to obtain

$$\begin{aligned} p_1(X) &= \int d\theta d\eta \; p(X|\theta,\eta)p(\theta,\eta) \\ p_0(X) &= \int d\theta d\eta \; p(X|\eta)p(\theta,\eta). \end{aligned} \tag{2-6}$$

Here θ and η refer to collections of parameters $\theta = \{\theta_i\}$, $\eta = \{\eta_i\}$. By the time the original signal waveform with parameters θ propagates to the various sensors over different paths, the aggregate parameter set θ gets expanded because some of the individual parameters (such as amplitude and phase) are modified along the way--requiring the assumption of different values in each channel--and perhaps some new parameters are acquired in the process (such as doppler shift). The situation becomes even more complex if the original parameters themselves are changing with time, so in this work we adopt the usual analytical procedure of considering the parameters constant over the interval of time they are being observed.

For another example in the way the processing task drives the data modelling effort, consider that, in order to locate the source, we should like to know the locations of the sensors. Not only that, but also conditioning and preprocessing parameters such as gains, sensor directivities and orientations, and the values of any time delays artificially introduced to the channels as a means of "beamforming." Then, too, we have to model the functional relationships between the source position and motion quantities and the quantities actually sensed. Therefore, the data being used for this task must be modelled to the extent of providing the required information _a priori_ or means for estimating it.

In this study we confine our scope to the tasks of detection and signal parameter extraction, with particular attention to performance of these tasks jointly. In all examples, we assume that the noise samples are jointly Gaussian and that signal and noise combine additively.

2.2 System Function Description

In Figure 2-1, the processor outputs were named somewhat cryptically as "decisions" and "numbers". Although decisions based upon inferences from the data could conceivably range from "drop a bomb" to "the source is a fish," we restrict ourselves to the classical choice between the hypotheses H_0 and H_1 concerning the presence of a signal. And, while what the "numbers" are will continue to be nebulous (in an effort to maintain generality), for the most part in the examples treated they will be estimates of parameters of the data waveforms, denoted by θ and η. Both the detection and estimation functions are to be carried out in ways that are optimal in some sense.

2.2.1 Optimal Detection

A review of the well known optimal detection procedures reveals the dependence of "optimality" upon the information postulated. Classically, the optimal detection problems is to find to the critical region Γ_1 such that if we

$$\text{decide:} \begin{cases} H_1 \text{ true if } X \text{ in } \Gamma_1 \\ H_0 \text{ true if } X \text{ in } \Gamma_0 = \Gamma - \Gamma_1 \end{cases} \tag{2-7}$$

then certain average "costs" are minimized. There are four possibilities connected with the decison, each in concept incurring a cost. The traditional notation is

true case	decision	cost
H_0	H_0	$C_{1-\alpha}$
H_0	H_1	C_α (type 1 error or false alarm)
H_1	H_0	C_β (type 2 error or miss)
H_1	H_1	$C_{1-\beta}$

By defining the error probabilities as

$$\alpha = Pr\{H_1 | H_0\} = \int_{\Gamma_1} dX p_0(X) \tag{2-8}$$

$$\beta = Pr\{H_0 | H_1\} = \int_{\Gamma_0} dX p_1(X) \tag{2-9}$$

and the probabilities of the occurrences themselves as

$$\pi_1 = 1 - \pi_0 = Pr\{H_1 \text{ true}\} , \tag{2-10}$$

then the average cost or "risk" is computed to be

$$R = \pi_0\left[\alpha C_\alpha + (1-\alpha)C_{1-\alpha}\right] + \pi_1\left[\beta C_\beta + (1-\beta)C_{1-\beta}\right] \tag{2-11}$$

$$= \int_{\Gamma_1} dX\left[\pi_0 C_\alpha p_0(X) + \pi_1 C_{1-\beta}\, p_1(X)\right]$$

$$+ \int_{\Gamma_0} dX\left[\pi_0 C_{1-\alpha} p_0(X) + \pi_1 C_\beta\, p_1(X)\right]$$

$$= \pi_0 C_\alpha + \pi_1 C_{1-\beta}$$

$$+ \int_{\Gamma_0} dX\left[\pi_1\left(C_\beta - C_{1-\beta}\right) p_1(X) - \pi_0\left(C_\alpha - C_{1-\alpha}\right) p_0(X)\right]. \tag{2-12}$$

For non-negative costs and situations in which mistakes are more costly than correct decisions, (2-12) is minimized if we pick Γ_0 such that

$$\Gamma_0 : \pi_1\left(C_\beta - C_{1-\beta}\right) p_1(X) < \pi_0\left(C_\alpha - C_{1-\alpha}\right) p_0(X) . \tag{2-13}$$

By defining a likelihood ratio

$$\Lambda(X) = p_1(X)/p_0(X) , \tag{2-14}$$

we can write the corresponding optimal decision rule as

$$\Lambda(X) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0\left(C_\alpha - C_{1-\alpha}\right)}{\pi_1\left(C_\beta - C_{1-\beta}\right)} , \tag{2-15}$$

or more generally as

$$\Lambda(X) \underset{H_2}{\overset{H_1}{\gtrless}} \lambda \tag{2-16}$$

where λ is a threshold. The procedure indicated by (2-15) is the Bayes decision rule, while (2-16) gives the Neyman-Pearson detector when λ is chosen to yield a fixed value of α (the resulting β is then minimum for that α). If $\lambda = 1$, then the "ideal observer" detector scheme is operative and the total probability of error is minimized.

In the foregoing, use of the marginal pdf's $p_0(X)$ and $p_1(X)$ includes the cases where these functions are computed from conditional and a priori pdf's, as shown in (2-6). Thus the likelihood ratio (2-14) may have the form

$$\Lambda(X) = \frac{\int d\theta d\eta p(X|\theta,\eta)p(\theta,\eta)}{\int d\theta d\eta p(X|\eta)p(\theta,\eta)} = \frac{E_{\theta,\eta}\{p(X|\theta,\eta)\}}{E_{\eta}\{p(X|\eta)\}} \tag{2-17}$$

When this is the case, it can be shown that the decision rule (2-15) is the one that also mimizies the a posteriori risk or average cost given the data.

What if the a priori probabilities π_0 and π_1 and/or pdf $p(\theta,\eta)$ are unknown? One procedure is to seek a "least favorable" (fictional) a priori distribution for the unknown parameters - one which maximizes the risk, - then minimize the risk as shown above to obtain a decision rule. This class of procedures, known as "minimax", works fine in principle, but often it is difficult to say what the least favorable distribution is. Moreover, the resulting decision rule is sometimes overly conservative, depending upon how "typical" the least favorable or "worst case" distribution is when actual data is processed. However, it can be stated that, if the risk based on the conditional pdf's does not depend upon the parameters, then the corresponding Bayes decision rule is also a minimax rule.

By far the most popular procedure to adopt when a priori information is missing is to ignore it and to use the conditional pdf's to compute the likelihood ratio after estimating the unknown parameters according to some criterion. One such criterion which is somewhat arbitrary with respect to detection risk, but has nice estimation properties, is the maximum likelihood criterion. Here, when ignorant of the exact distribution of the parameters θ and η, we select values for them which make the observed data X most likely; that is θ_1, η_1, and η_0 are chosen such that

$$\begin{aligned} p(X|\theta_1,\eta_1) &\geq p(X|\theta,\eta) \\ p(X|\eta_0) &\geq p(X|\eta) \ . \end{aligned} \qquad (2\text{-}18)$$

(often $\eta_0 \equiv \eta_1$). The resulting decision rule, while retaining the Bayes form -- that is, likelihood ratio detection -- does not necessarily involve minimum risk or even an "acceptable" risk. However, if the data base is satisfactory, then the quality of the estimates and also the decision will be acceptable.

2.2.2 Optimal Estimation

Reviewing established parameter estimation procedures also shows that optimality criteria can be influenced by the amount of a priori information available.

Basic properties of estimators are summarized in Table 2-1, in which for simplicity we speak of the observed data X and one parameter set θ. These basic properties, shown for both the case of conditional

TABLE 2-1

PROPERTIES OF ESTIMATORS

	CONDITIONAL	UNCONDITIONAL
bias	$b(\theta) = E_{X\|\theta}\{\hat{\theta}(X)\} - \theta$ $\equiv E\{\hat{\theta}-\theta\|\theta\}$	$B = E_{X,\theta}\{\hat{\theta}(X)\} - E_{\theta}\{\theta\}$ $= E_{\theta}\{b(\theta)\}$ $\equiv E\{\hat{\theta}-\theta\}$
mean square error (MSE)	$e^2(\theta) = E_{X\|\theta}\left\{\left[\hat{\theta}(X)-\theta\right]^2\right\}$ $\equiv E\left\{(\hat{\theta}-\theta)^2\|\theta\right\}$	$e^2 = E_{X,\theta}\left\{\left[\hat{\theta}(X)-\theta\right]^2\right\} = E\{e^2(\theta)\}$ $\equiv E\left\{(\hat{\theta}-\theta)^2\right\}$
efficiency	$\varepsilon(\theta) = \dfrac{e^2(\theta)\min}{e^2(\theta)} \leq 1$	$\varepsilon = e^2\min/e^2 \leq 1$
minimum MSE	$\dfrac{\left[1+b'(\theta)\right]^2}{E\left\{\left[\frac{\partial}{\partial\theta}\ln p(X\|\theta)\right]^2\|\theta\right\}}$	$\dfrac{E\left\{\left[1+b'(\theta)\right]^2\right\}}{E\left\{\left[\frac{\partial}{\partial\theta}\ln p(X,\theta)\right]^2\right\}}$
unbiased	$b(\theta) = 0$	$B = 0$
efficient	$\varepsilon(\theta) = 1$	$\varepsilon = 1$
sufficient	$p(\theta\|X) = p\left[\theta\|\hat{\theta}(X)\right]$ or $p(\theta\|X) = f(X)g\left[\hat{\theta}(X),\theta\right]$	$p(X) = f(X)E_{\theta}\left\{g\left[\hat{\theta}(X),\theta\right]\right\}$

estimates (based on $p(X|\theta)$) and unconditional (based on $p(X,\theta)$), are often extended to define more subtle properties, such as asymptotic efficiency, when sample size and various convergence criteria are taken into account. In this brief review we restrict ourselves to the basics and consider two large classes of estimators, Bayes and Maximum likelihood (ML).

In a manner very similar to that shown above for the Bayesian detector, when a priori parameter pdf's are available, we can construct estimators $\hat{\theta}(X)$ of the unknown parameters θ from the data which minimize the risk or average cost of estimation. In this case we write the risk as

$$R = E_{X,\theta}\{C[\theta,\hat{\theta}(X)]\}$$
$$= \int dx d\theta \, C[\theta,\hat{\theta}(X)] \, p(X,\theta). \qquad (2\text{-}19)$$

The necessary minimization with respect to $\hat{\theta}$ requires specifying the cost function to some degree; also the form of the estimator thus obtained depends upon the cost function. Therefore when talking unconditional estimators, one has to refer to the specific class of cost function for which the estimator is optimal.

For the so-called "simple cost function" (SCF), we use

$$C(\theta-\hat{\theta}) = C_e-(C_e- C_c)\delta(\hat{\theta}-\theta) \qquad (2\text{-}20a)$$

or

$$C(\theta-\hat{\theta}) = \begin{cases} C_e, & \hat{\theta} \neq \theta \\ C_c, & \hat{\theta} = \theta \, . \end{cases} \qquad (2\text{-}21)$$

Using the first notation, it is easy to see that the optimal estimator in the case of the SCF is the unconditional ML estimator, given by

$$p(X,\hat{\theta}) \geq p(X,\theta). \qquad (2\text{-}21)$$

For the fidelity criteria, or quadratic cost function (QCF),

$$C(\theta-\hat{\theta}) = C_0(\hat{\theta}-\theta)^2 \tag{2-22}$$

the resulting "Bayes estimator" is

$$\hat{\theta} = E\{\theta|X\}, \tag{2-23}$$

the *a posteriori* mean of the parameter. Statements that can be made about the Bayes estimator for the QCF are the following:

(a) it is conditionally unbiased: $b(\theta) = 0$

(b) it has the smallest average variance among all unbiased estimators

(c) if the joint distribution $p(X,\theta)$ is unimodal and symmetric about the mode (i.e., mean = mode) with respect to θ, then the Bayes estimator is the maximum *a posteriori* estimate.

(d) it is the optimal estimator if the cost function (other than QCF) is even about $\hat{\theta} = \theta$ and if the *a posteriori* pdf $p(\theta|X)$ has mean = mode.

With only the conditional pdf $p(X|\theta)$ available, we use the (conditional) ML estimator,

$$p(X|\hat{\theta}) \geq p(X|\theta) \ . \tag{2-24}$$

Although in general ML estimates are biased and not unique, they have good asymptotic features (under rather general assumptions) such as consistency (asymptotic unbiasedness) and efficiency. Also, ML estimators are sufficient statistics (or functions of them) when they exist, and when efficient estimators exist, they are ML estimators. On the whole, then, unless cost functions are an integral aspect of the problem, people find ML estimation to be the convenient route to follow.

2.3 Procedures adopted for this study

Having reviewed very briefly some classical results from detection and estimation theory for reference, we proceed in the light of our stated objectives in the following way.

In Chapter 3 we pursue the notion that optimal joint detection and estimation involves detectors and estimators that are possibly different from those obtained separately, using two basic approaches to the task, and developing single-channel examples to illustrate them.

In Chapter 4 the detection and estimation tasks are formulated for data in the form of matrices in an attempt to discover in what ways processors designed to operate on m channels simultaneously differ from combinations of single-channel operations.

3.0 JOINT DETECTION AND ESTIMATION

Intuitively it seems reasonable that an optimal procedure for jointly performing detection of a signal in noise and estimation of parameters would involve a processor structure in general different from a simple combination of the optimal procedures for performing these tasks separately. One might also anticipate that the joint operation would in some sense be better as well as different - perhaps more efficient if not more accurate.

Optimality, of course, is with respect to given criteria. "Joint operation" also is a nebulous concept without further definition. In this chapter two approaches to joint optimal detection and estimation are presented and amplified. One is quite systematic and unified, using cost functions to "couple" the two operations (after Middleton and Esposito). The second is more *ad hoc* in nature, exploiting the common use by both operations of sufficient statistics to achieve economical computation (after Birdsall and Gobien).

3.1 Cost Coupled Detection and Estimation

In addition to the notations already introduced, we define the cost functions

$C_{00}(\hat{\eta};\eta)$ = cost of accepting H_0 and estimating noise parameters when H_0 is true

$C_{10}(\hat{\theta},\hat{\eta};\eta)$ = cost of accepting H_1 and estimating signal and noise parameters when H_0 is true

$C_{01}(\hat{\eta};\theta,\eta)$ = cost of accepting H_0 and estimating only noise parameters when H_1 is true

$C_{11}(\hat{\theta},\hat{\eta};\theta,\eta)$ = cost of accepting H_1 and estimating signal and noise parameters when H_1 is true

With these costs defined, we can calculate an unconditional average cost or risk as

$$R = E\{r(\theta,\eta)\} \tag{3-1}$$

where the conditional risk is given by

$$r(\theta,\eta) = \int_{\Gamma_0} dX \left[\pi_0\, C_{00}(\hat{\eta};\eta)\, p(X|\eta) + \pi_1\, C_{01}(\hat{\eta};\theta,\eta)\, p(X|\theta,\eta)\right]$$

$$+ \int_{\Gamma_1} dX \left[\pi_0\, C_{10}(\hat{\theta},\hat{\eta};\eta)\, p(X|\eta) + \pi_1\, C_{11}(\hat{\theta},\hat{\eta};\theta,\eta)\, p(X|\theta,\eta)\right] . \tag{3-2}$$

We now wish to minimize R with respect to both the critical region Γ_1 and the estimators $\hat{\theta}(X)$ and $\hat{\eta}(X)$. This is accomplished in two steps: first, with respect to Γ_1, then $\hat{\theta}$ and $\hat{\eta}$.

The conditional risk can be rewritten as

$$r(\theta,\eta) = \int_{\Gamma_0} dX\, A(X,\hat{\eta};\theta,\eta) + \int_{\Gamma_1} dX\, B(X,\hat{\theta},\hat{\eta};\theta,\eta)$$

$$= \int_{\Gamma_0+\Gamma_1} dX\, A + \int_{\Gamma_1} dX\, (B-A) , \tag{3-3}$$

in which the identification of A and B is discerned from (3-2). Granting that the functions A and B are everywhere positive, r is minimized if Γ_0 and Γ_1 are selected so that

$$\Gamma_1 : B<A,\ \Gamma_0 : B\geq A . \tag{3-4}$$

With the estimators yet to be specified, the corresponding conditional and unconditional decision rules are

$$\Lambda_g(X;\theta,\eta) = \frac{(C_{01}-C_{11})\, p(X|\theta,\eta)}{(C_{10}-C_{00})\, p(X|\eta)} \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0/\pi_1 \tag{3-5}$$

and

$$\Lambda_g(X) = \frac{E_{\theta,\eta}\left\{(C_{01}-C_{11})\, p(X|\theta,\eta)\right\}}{E_\eta\left\{(C_{10}-C_{00})\, p(X|\eta)\right\}} \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0/\pi_1 \,, \tag{3-6}$$

with Λ_g denoting "generalized likelihood ratio."

3.1.1 Differentiable cost functions

Minimization with respect to the estimators $\hat{\theta}$ and $\hat{\eta}$ is carried out by differentiating R inside the integrals, resulting in the simultaneous constraints

$$E_{\theta,\eta}\left\{\frac{\partial A}{\partial \hat{\eta}}\right\} = E_{\theta,\eta}\left\{\frac{\partial B}{\partial \hat{\eta}}\right\} = E_{\theta,\eta}\left\{\frac{\partial B}{\partial \hat{\eta}}\right\} = 0 \tag{3-7}$$

Looking at these requirements carefully, we find that they are equivalent to

$$E_{\theta,\eta}\left\{\pi_0 \frac{\partial C_{00}}{\partial \hat{\eta}} p(X|\eta) + \pi_1 \frac{\partial C_{10}}{\partial \hat{\eta}} p(X|\theta,\eta)\right\} = 0 \tag{3-8a}$$

$$E_{\theta,\eta}\left\{\pi_0 \frac{\partial C_{10}}{\partial \hat{\eta}} p(X|\eta) + \pi_1 \frac{\partial C_{11}}{\partial \hat{\eta}} p(X|\theta,\eta)\right\} = 0 \tag{3-8b}$$

and

$$E_{\theta,\eta}\left\{\pi_0 \frac{\partial C_{10}}{\partial \hat{\eta}} p(X|\eta) + \pi_1 \frac{\partial C_{11}}{\partial \hat{\theta}} p(X|\theta,\eta)\right\} = 0 \tag{3-8c}$$

Since noise (and noise parameters) are present under acceptance of hypothesis, it is reasonable to assume that

$$\frac{\partial C_{00}}{\partial \hat{\eta}} = \frac{\partial C_{10}}{\partial \hat{\eta}} \quad \text{and} \quad \frac{\partial C_{01}}{\partial \hat{\eta}} = \frac{\partial C_{11}}{\partial \hat{\eta}} \tag{3-9}$$

so that (3-8a) and (3-8b) are the same equation. This presupposes, actually, that the cost of estimating the noise parameters η is additively

combined to the other cost components and with the same weight under either decision. If we further presuppose that the same weight holds under the truth of either hypothesis, we arrive at

$$E_{\eta}\left\{\frac{\partial C}{\partial \hat{\eta}}\left[\pi_0 p(X|\eta) + \pi_1 E_{\theta}\{p(X|\theta,\eta)\}\right]\right\} = 0 \tag{3-10}$$

For example, a set of cost functions based on the QCF might be

$$\begin{aligned}
C_{00} &= C_{1-\alpha} + C_N(\hat{\eta}-\eta)^2 \\
C_{10} &= C_{\alpha} + C_N(\hat{\eta}-\eta)^2 + C_a\theta^2 \\
C_{01} &= C_{\beta} + C_N(\hat{\eta}-\eta)^2 + C_b\theta^2 \\
C_{11} &= C_{1-\beta} + C_N(\hat{\eta}-\eta)^2 + C_s(\hat{\theta}-\theta)^2 \,.
\end{aligned} \tag{3-11}$$

These costs lead to the estimates

$$\hat{\eta} = E\{\eta|X\} = \pi_0\, E\left\{\eta|X,H_0\right\} + \pi_1\, E\left\{\eta|X,H_1\right\} \tag{3-12}$$

$$\hat{\theta} = \frac{\Lambda'}{1+\Lambda'}\, E\left\{\theta|X,H_1\right\},\ \Lambda' = \frac{\pi_1\, C_s}{\pi_0\, C_a}\Lambda, \tag{3-13}$$

and the (unconditional) decision rule

$$\Lambda_g(X) = \frac{C_{\beta}-C_{1-\beta} + E\left\{C_b\theta^2 - C_s(\hat{\theta}-\theta)^2|X\right\}}{C_{\alpha}-C_{1-\alpha} + C_a\hat{\theta}^2(X)}\,\Lambda(X) \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0/\pi_1 \tag{3-14}$$

or

$$\Lambda(X)\left[1 + \frac{C_s\hat{\theta}^2(X)}{C_{\beta}-C_{1-\beta}} + \frac{\left(C_b-C_s\right)E\left\{\theta^2|X\right\}}{C_{\beta}-C_{1-\beta}}\right] + \frac{\pi_0}{\pi_1}\,\frac{C_a\hat{\theta}^2(X)}{C_{\beta}-C_{1-\beta}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0\left(C_{\alpha}-C_{1-\alpha}\right)}{\pi_1\left(C_{\beta}-C_{1-\beta}\right)}\,. \tag{3-15}$$

Immediately we see that, under this cost assignment, the noise parameter estimate is the same as in separate D/E, the signal parameter estimate is modified by the separate D/E likelihood ratio, and the decision rule is considerably more complicated.

EXAMPLE: Suppose a sinewave is to be detected in Gaussian noise. The appropriate conditional pdf's may be written

$$p(X|\alpha,\beta,N) = \left(\frac{1}{2\pi N}\right)^n \exp\left\{-\frac{1}{2N}\sum_{i=1}^{n}\left[\left(u_i-\alpha\right)^2 + \left(v_i-\beta\right)^2\right]\right\} \tag{3-16}$$

and

$$p(X|N) = \left(\frac{1}{2\pi N}\right)^n \exp\left\{-\frac{1}{2N}\sum_{i=1}^{n}\left(u_i^2 + v_i^2\right)\right\}. \tag{3-17}$$

The parameters α, β and N are considered to be independent, with a priori pdf's

$$p(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\alpha-\alpha_0}{\sigma}\right)^2\right\} \tag{3-18}$$

$$p(\beta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\beta-\beta_0}{\sigma}\right)^2\right\} \tag{3-19}$$

$$p(N) = \left[2\gamma_0\lambda_0 K_1\left(2\lambda_0\right)\right]^{-1} \exp\left\{-\frac{N}{\gamma_0} - \frac{\gamma_0\lambda_0^2}{N}\right\}, \quad N > 0. \tag{3-20}$$

It can be shown that in this case

$$p_0(X) = k_n \left[\frac{1}{2}\gamma_0 P_n + \lambda_0^2\gamma_0^2\right]^{(1-n)/2} K_{n-1}\left(2\sqrt{\lambda_0^2 + P_n/2\gamma_0}\right) \tag{3-21}$$

and

$$p_1(X) \approx k_n \left(\frac{1}{n\sigma^2}\right) \left[\frac{1}{2}\gamma_0 R_n + \lambda_0^2\gamma_0^2\right]^{1-n/2} K_{n-2}\left(2\sqrt{\lambda_0^2 + R_n/2\gamma_0}\right), \quad N \ll n\sigma^2 \tag{3-22}$$

with

$$P_n = \sum_{i=1}^{n}\left(u_i^2 + v_i^2\right) \tag{3-23a}$$

$$R_n = P_n - \frac{1}{n}\left[\left(\sum u_i\right)^2 + \left(\sum v_i\right)^2\right] \tag{3-23b}$$

and

$$k_n = \left[(2\pi)^n \lambda_0\gamma_0 K_1\left(2\lambda_0\right)\right]^{-1}. \tag{3-23c}$$

The *a posteriori* means are

$$E\left\{N|X,H_0\right\} = \gamma_0\sqrt{\lambda_0^2 + P_n/2\gamma_0}\;\frac{K_{n-2}}{K_{n-1}}\left\{2\sqrt{\lambda_0^2 + P_n/2\gamma_0}\right\} \tag{3-24a}$$

$$E\left\{N|X,H_1\right\} = \gamma_0\sqrt{\lambda_0^2 + R_n/2\gamma_0}\;\frac{K_{n-3}}{K_{n-2}}\left\{2\sqrt{\lambda_0^2 + R_n/2\gamma_0}\right\} \tag{3-24b}$$

$$E(\alpha|X) = E\left\{\frac{\alpha_0 N + n\sigma^2\,\bar{u}}{N+n\sigma^2}\,|X,H_1\right\} \tag{3-25}$$

$$= \bar{u} + \frac{\alpha_0}{n\sigma^2}\;E\left\{N|X,H_1\right\}$$

$$E(\beta|X) = E\left\{\frac{\beta_0 N + n\sigma^2\,\bar{v}}{N + n\sigma^2}\,|X,H_1\right\}$$

$$= \bar{v} + \frac{\beta_0}{n\sigma^2}\;E\left\{N|X,H_1\right\} \tag{3-26}$$

We note that these expressions involve the statistics P_n, R_n, $\bar{u}$, and $\bar{v}$, which are the sufficient statistics in this case. Having found these expressions, then, we can diagram the QCF - coupled D/E processor as shown in Figure 3-1.

3.1.2 Simple Cost Function (SCF)

For the SCF, minimization of R for coupled D/E cannot be carried out by differentiation of the cost functions since they take the form

$$C_{00} = C_{1-\alpha} + C_e - \Delta\delta(\hat{\eta}-\eta)\delta(\hat{\theta})$$
$$C_{10} = C_{\alpha} + C_e - \Delta\delta(\hat{\eta}-\eta)\delta(\hat{\theta})$$
$$C_{01} = C_{\beta} + C_e - \Delta\delta(\hat{\eta}-\eta)\delta(\hat{\theta}-\theta)$$
$$C_{11} = C_{1-\beta} + C_e - \Delta\delta(\hat{\eta}-\eta)\delta(\hat{\theta}-\theta) \tag{3-27}$$

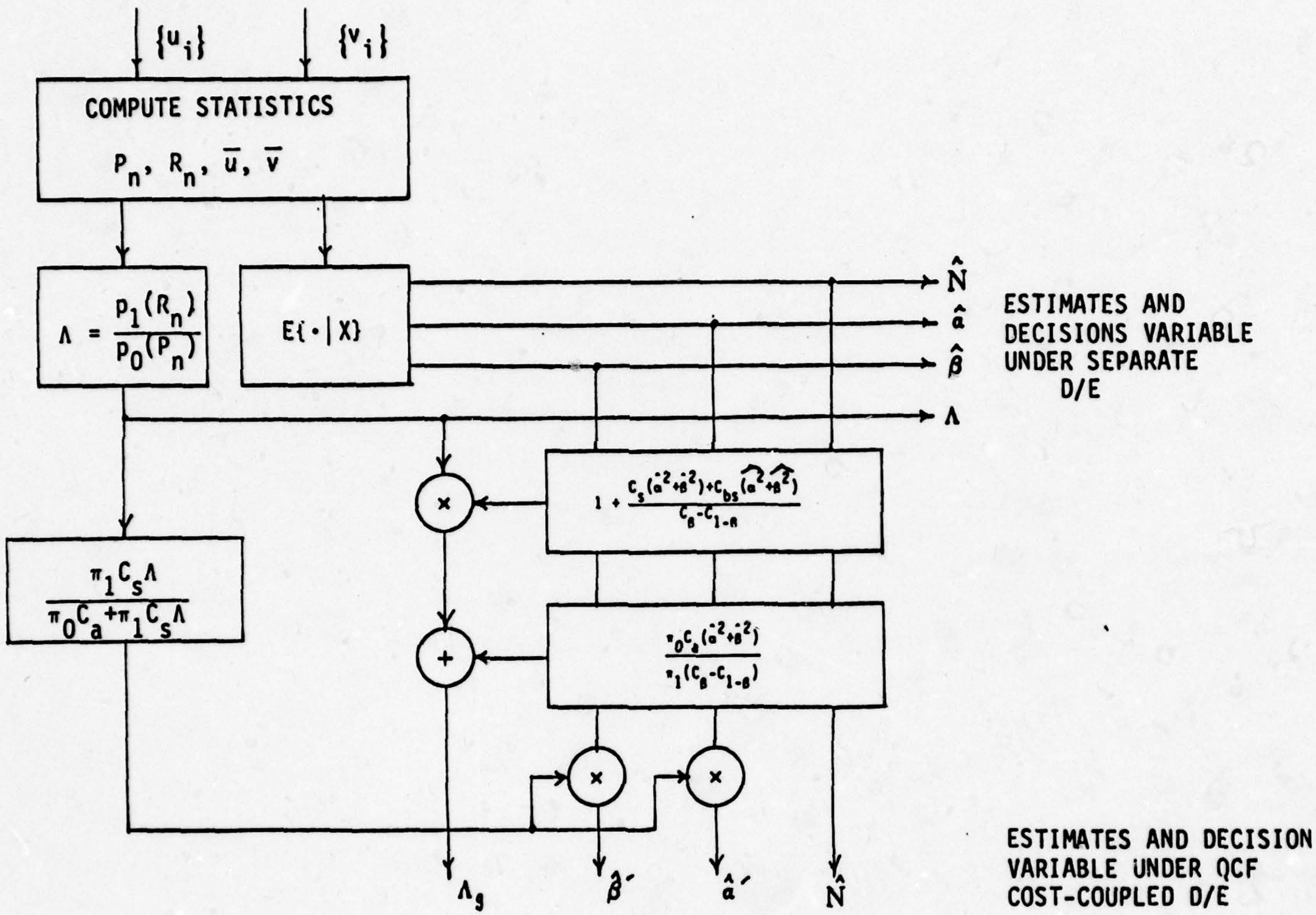


FIGURE 3-1 COST-COUPLED DETECTION AND ESTIMATION SYSTEM

To these costs there corresponds the risk

$$R = C_e + \pi_0 \left[C_{1-\alpha}(1-\alpha) + C_\alpha \alpha\right] + \pi_1 \left[C_{1-\beta}(1-\beta) + C_\beta \beta\right]$$

$$- \pi_0 \Delta \int_{\Gamma_0} dX\, p_0(X,\hat{\eta})\delta(\hat{\theta}) - \pi_1 \Delta \int_{\Gamma_0} dX\, p_1(X,\hat{\theta},\hat{\eta})$$

$$- \pi_0 \Delta \int_{\Gamma_1} dX\, p_0(X,\hat{\eta})\delta(\hat{\theta}) - \pi_1 \Delta \int_{\Gamma_1} dX\, p_1(X,\hat{\theta},\hat{\eta})$$

$$= C_e + R_D - \Delta \int_{\Gamma} dX \left\{\pi_0 \delta(\hat{\theta}) p_0(X,\hat{\eta}) + \pi_1 p_1(X,\hat{\theta},\hat{\eta})\right\}, \qquad (3\text{-}28)$$

in which it is understood that $\hat{\eta}=\hat{\eta}(X)$ and $\hat{\theta}=\hat{\theta}(X)$. The risk is minimum when the integral is maximum:

$$\pi_1 \int_{\Gamma} dX\, p_1(X,\hat{\theta},\hat{\eta}) + \pi_0 \int_{\Gamma-\Gamma_{\hat{\theta}}} dX\, p_0(X,\hat{\eta})$$

$$\geq \pi_1 \int_{\Gamma} dX\, p_1(X,\theta,\eta) + \pi_0 \int_{\Gamma-\Gamma_\theta} dX\, p_0(X,\eta); \qquad (3\text{-}29)$$

using

$$X \epsilon \Gamma_\theta \text{ when } \theta(x)=0. \qquad (3\text{-}30)$$

Thus it is evident that the estimators $\hat{\theta}$ and $\hat{\eta}$ in this case are not unconditional ML estimators as in the separate D/E situation, but can be viewed as a kind of generalization of them.

3.2 Combined Adaptive D/E

One criticism which is often made against the Bayesian approach in general is that the formal, optimal solution using the approach requires information that may not be available in practical situations, such as *a priori* distributions for unknown parameters. In these situations one can simply guess a likely distribution or other needed specification and proceed with the Bayesian derivation, not knowing whether the resulting design will be satisfactory (mostly likely, if intelligent guesses were made, the system will work but we do not expect it to perform like a true optimal system). If guessing is too risky or unesthetic, then minimax or maximum likelihood approaches can be used, corresponding respectively to "worst case" or conditional optimization as discussed in Chapter 2.

In this section we consider aspects of another, adaptive approach. Sometimes called the "empirical Bayes approach," the object is to take advantage of certain properties of probability distributions to obtain in effect, an estimate of the required *a priori* information as data are being taken. Robbins has shown that it is possible to construct decision functions (detectors and estimators), with respect to the distributions of unknown parameters, which asymptotically (as more and more observations conditioned on the unknown parameters are taken) incur the minimum or Bayes risk.

Spragins discusses "reproducing" properties of probability distributions, noting that the *a posteriori* pdf for parameters θ on data $\{x_i\}$ has the iterative formulation

$$p(\theta|X_n) = \frac{p(X_n|\theta)p(\theta)}{\int d\theta(\text{numerator})}$$

$$= \frac{p(\theta|X_{n-1})p(x_n|\theta)}{\int d\theta(\text{numerator})}, \tag{3-31}$$

where $X_n \equiv (x_1, x_2, \ldots, x_n)$ represents n independent samples of the data. In particular, as the number of samples becomes large, the a posteriori pdf eventually approaches a narrow spike centered around the true value of θ, provided the original (a priori) pdf $p(\theta)$ is defined on an interval containing the true value of θ.

A simple illustration of this principle is the following: suppose samples from a Gaussian population with unit mean and variance are observed. However, the true mean is not known, so the following triangular a priori pdf is postulated:

$$p(\theta) = \begin{cases} 1 - |\theta - \frac{1}{2}| , & |\theta - \frac{1}{2}| \leq 1 \\ 0, & \text{elsewhere} \end{cases} . \tag{3-32}$$

After n observations, the a posteriori is proportional to

$$p(\theta|X_n) = K \exp\left\{-\frac{1}{2}\sum_i (x_i-\theta)^2\right\} p(\theta)$$

$$= K \exp\left\{-\frac{n}{2}(\theta-\bar{x})^2 - \frac{1}{2}\sum(x_i-\bar{x})^2\right\} p(\theta)$$

$$= K' \exp\left\{-\frac{n}{2}(\theta-1)^2\right\}\left[1-|\theta-\tfrac{1}{2}|\right], \quad |\theta-\tfrac{1}{2}| \leq 1. \tag{3-33}$$

As demonstrated in Figure 3-2, the peak of $p(\theta|X_n)$ does indeed approach the correct value.

What this phenomenon suggests is that we can select a priori pdf's---that is, functional forms---which "look like" the a posteriori pdf after a number of fictional prior data observations. We then can

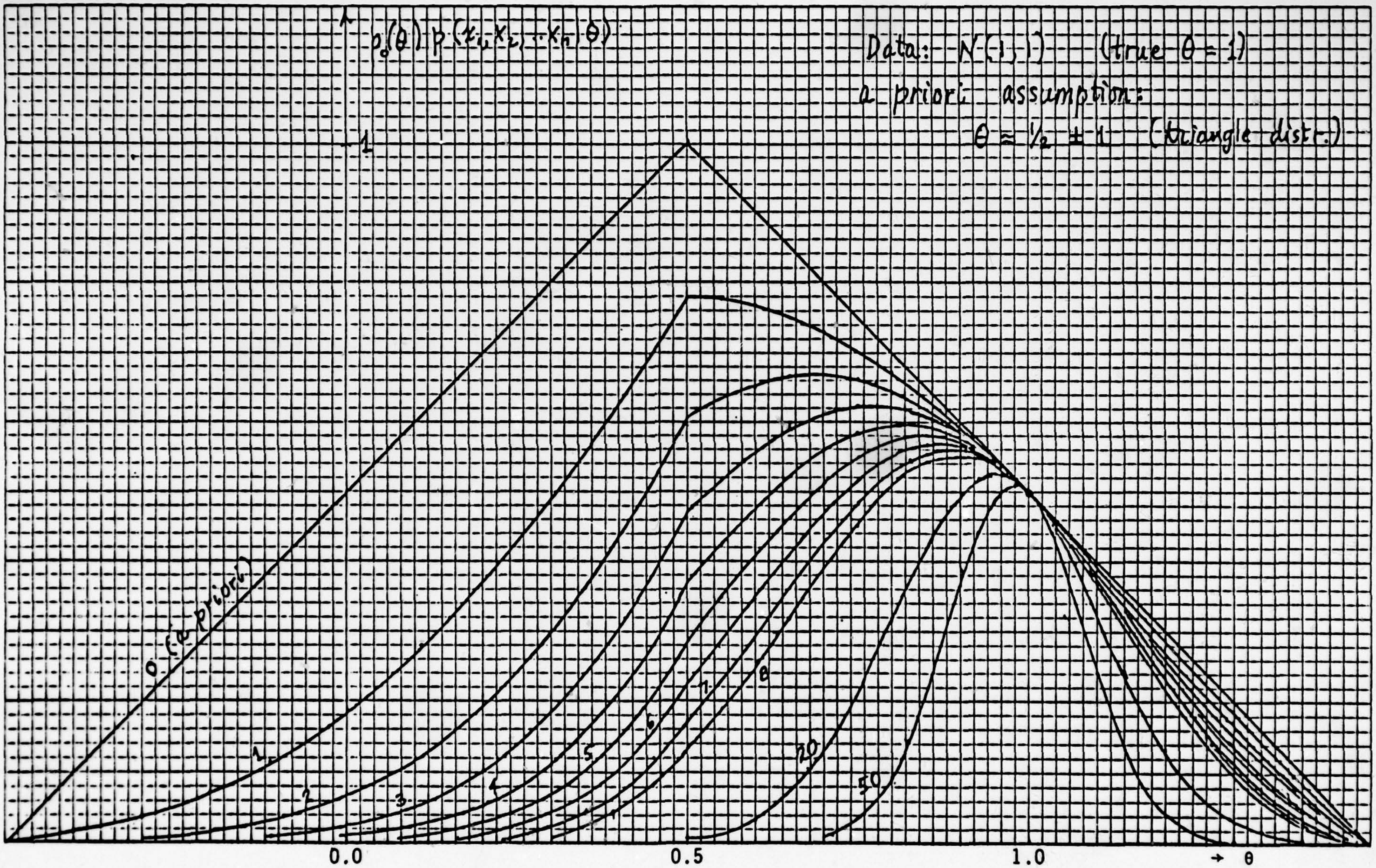


FIGURE 3-2 ITERATIVE APPLICATION OF BAYES' RULE

expect the same functional form to be preserved, its parameters taking on values which depend on the data and which converge eventually to the correct values. For example, in the previous illustration we would choose

$$p(\theta) = K \exp\left\{- \frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma_0}\right)^2\right\} \qquad (3\text{-}34)$$

θ_0 being an initial guess of the true value for θ. The data then generates

$$p(\theta|X_n) = K \exp\left\{- \frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma}\right)^2 - \frac{1}{2}\sum\left(\frac{x_i-\theta}{\sigma}\right)^2\right\}$$

$$= K' \exp\left\{- \frac{1}{2}\left(\frac{\theta-\theta_n}{\sigma_n}\right)^2\right\}$$

with

$$\theta_n = \frac{\sigma^2\theta_0 + \sigma_0^2\sum x_i}{\sigma^2 + n\,\sigma_0^2} \approx \frac{1}{n}\sum x_i = \bar{x}, \text{ large } n. \qquad (3\text{-}35)$$

Another concern of a computational nature is that the a posteriori pdf not only reproduce in this sense, but also that the parameters specifying or "indexing" the distribution remain fixed in dimension. This property is insured if the parameters in question have corresponding to them sufficient statistics in the data.

These concepts can be used to real advantage in combined detection and estimation, as shown by Birdsall and Gobien. The (marginal) likelihood function can be developed in the following way to yield an iterative or sequential form:

$$L(X_n) \equiv \frac{p(X_n|H_1)}{p(X_n|H_0)}$$

$$= \frac{p(\theta,\eta|H_1)}{p(\theta,\eta|X_n,H_1)} \cdot \frac{p(\eta|X_n,H_0)}{p(\eta|H_0)} \cdot L(X_n|\theta,\eta) \qquad (3\text{-}36)$$

by Bayes' rule, where $L(X_n|\theta,\eta)$ is the known parámeter (conditional) likelihood function. This may be written also

$$L(X_n) = \prod_{i=1}^{n} L(x_i|X_{i-1})$$

$$= \prod_{i=1}^{n} \frac{p(x_i|X_{i-1},H_1)}{p(x_i|X_{i-1},H_0)}$$

$$= \prod_{i=1}^{n} \frac{p(\theta,\eta|X_{i-1},H_1)}{p(\theta,\eta|X_i,H_1)} \cdot \frac{p(\eta|X_i,H_0)}{p(\eta|X_{i-1},H_0)} L(x_i|X_{i-1},\theta,\eta). \qquad (3\text{-}37)$$

Note especially that $L(X_n)$ is not only now in a form for iterative computation, but also that the choice of θ and η to evaluate the expression is arbitrary, leaving the designer free to pick what is convenient.

EXAMPLE: A random sample is taken from a Gaussian population whose mean and variance are unknown. The hypothesis H_0: $\mu=0$ is to be tested against the alternative H_1: $\mu\neq 0$. According to our present formulation we find for H_1

$$p(\theta,\eta|X_n) = \frac{p(X_n|\theta,\eta)p(\theta,\eta)}{\int d\theta d\eta(\text{numerator}}$$

$$= K\eta^{-n/2} \exp\left\{-\frac{1}{2\eta}\sum(x_i-\theta)^2\right\} p(\theta,\eta)$$

$$= K\eta^{-n/2} \exp\left\{-\frac{1}{2\eta}\left[n(\theta-\bar{x})^2 + \sum(x_i-\bar{x})^2\right]\right\} p(\theta,\eta), \qquad (3\text{-}38)$$

where θ stands in the place of the unknown mean and η, the unknown variance. This form suggests choosing

$$p(\theta,\eta) = K\eta^{-n_0/2} \exp\left\{-\frac{n_0}{2\eta}\left[(\theta-\theta_0)^2 + \eta_0\right]\right\} \tag{3-39}$$

for which the modes are θ_0 and η_0.

The resulting a posteriori pdf is then reproducing:

$$p(\theta,\eta|X_n) = a_n\eta^{-\nu/2} \exp\left\{-\frac{\nu}{2\eta}\left[(\theta-\theta_n)^2 + \eta_n\right]\right\}, \tag{3-40}$$

with $\nu = n + n_0$ and

$$\theta_n = \frac{n_0\,\theta_0 + n\overline{x}_n}{n_0 + n} \tag{3-41a}$$

$$\eta_n = \frac{n_0\eta_0 + \sum(x_i-\overline{x}_n)^2}{n_0 + n} + \frac{nn_0(\overline{x}_n-\theta_0)^2}{(n + n_0)^2} \tag{3-41b}$$

$$a_n = \left[\sqrt{\frac{2\pi}{\nu}}\left(\frac{\nu\eta_n}{2}\right)^{(\nu-3)/2} \Gamma\left(\frac{\nu-3}{2}\right)\right]^{-1}. \tag{3-41c}$$

For H_0 we have, analogous to (3-38),

$$p(\eta|X_n) = K\eta^{-n/2} \exp\left\{-\sum x_i^2/2\eta\right\} p(\eta), \tag{3-42}$$

suggesting that we choose the a priori pdf

$$p(\eta) = K\eta^{-n_0/2} \exp\left\{-n_o\lambda_o/2\eta\right\} \tag{3-43}$$

with the mode λ_0. The a posteriori pdf then is

$$p(\eta|X_n) = b_n\eta^{-\nu/2} \exp\left\{-\frac{\nu\lambda_n}{2\eta}\right\} \tag{3-44}$$

with $\nu = n_o + n$ again and

$$\lambda_n = \frac{n_0\lambda_0 + \sum x_i^2}{n_0 + n} \tag{3-45a}$$

$$b_n = \left[\left(\frac{\nu\lambda_n}{2}\right)^{(\nu-2)/2} \Gamma\left(\frac{\nu-2}{2}\right)\right]^{-1}. \tag{3-45b}$$

The iterative form for the likelihood function can now be specified as

$$L(x_n|X_{n-1}) = \frac{a_{n-1}b_n}{a_n b_{n-1}} \exp\left\{\frac{1}{2\eta}\left[(\theta-\theta_{n-1})^2 + \eta_{n-1} - \lambda_{n-1}\right]\right.$$

$$\left. - \frac{\nu}{2\eta}\left[(\theta_{n-1}+\theta_n - 2\theta)(\theta_{n-1}-\theta_n) + \eta_{n-1}-\lambda_{n-1}-\eta_n + \lambda_n\right]\right\}$$

$$\times \quad L(x_n|X_{n-1},\theta,\eta)$$

$$= \frac{a_{n-1}b_n}{a_n b_{n-1}} \exp\left\{-\frac{1}{2\eta}\left[\nu(\theta_{n-1} + \theta_n - 2\theta)(\theta_{n-1} - \theta_n) - (\theta-\theta_{n-1})^2\right.\right.$$

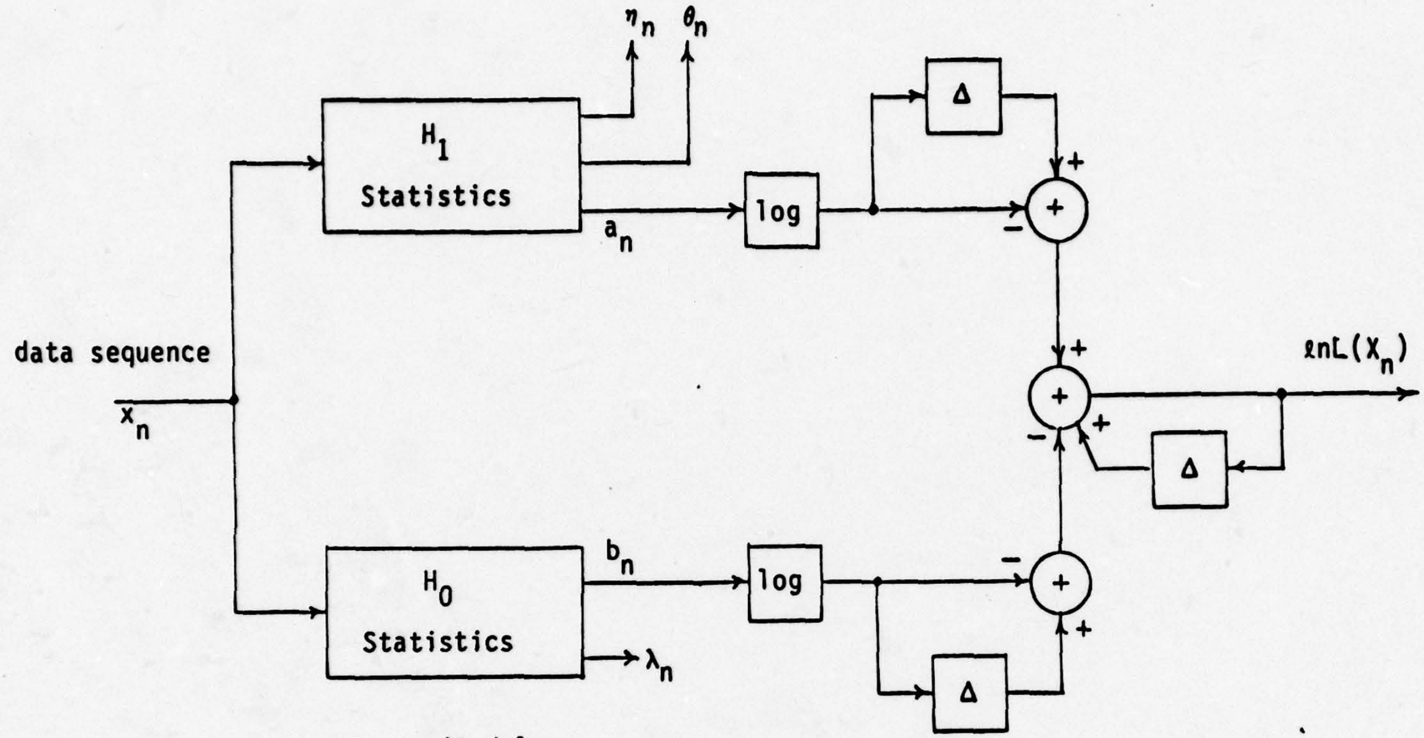
$$\left.\left. + \theta(\theta-2x_n) + n\bar{x}_n^2 - (n-1)\bar{x}_{n-1}^2 + \varepsilon\right]\right\} \tag{3-46}$$

Since, from (3-37) and comments following, we may use any admissible fixed values of θ and η to evaluate the likelihood function, we can base the detection on $\eta \to \infty$:

$$L(x_n|X_{n-1}) = \frac{a_{n-1}b_n}{a_n b_{n-1}}$$

$$= \sqrt{\frac{\nu-1}{\nu}}\; \frac{B\left(\frac{1}{2}, \frac{\nu-3}{2}\right)}{B\left(\frac{1}{2}, \frac{\nu-4}{2}\right)}\; \frac{\left[\nu(\nu-1)\;\eta_n\lambda_{n-1}\right]^{(\nu-3)/2}}{\left[(\nu-1)\eta_{n-1}\right]^{(\nu-4)/2}\left[\nu\lambda_n\right]^{(\nu-2)/2}} \tag{3-47}$$

Using $z_n = \ell n\, L(x_n|X_{n-1})$, we can implement detection as shown in Figure 3-3.



Δ = unit delay

FIGURE 3-3 JOINT SEQUENTIAL DETECTION AND ESTIMATION

4.0 MULTIPLE SENSOR APPLICATIONS

What happens to the detection and estimation (D/E) structures discussed in the previous chapters, when the data is a vector or matrix? Or, what if the signal and/or noise parameters are elements of a vector or a matrix? These questions are related to the tópic of this chapter, which is to extend the D/E results to the situation in which data from more than one channel or sensor are received and it is desired to process them in a way that takes into consideration the inter-channel dependencies that exist in general.

The techniques of the previous two chapters can be, and have been, developed to a high degree of effectiveness by statisticans and, to a lesser extent, by engineers. So far we have seen that optimal joint detection and estimation procedures tend to involve the usual Bayes parameter estimates (or slight modifications to them) and fairly complicated detectors. We have also seen that an empirical Bayes procedure can be used to perform asymptotically optimum detection with sequentially learned parameter pdf's as a by-product. The existence of these techniques indicates that the processing system of Figure 1-1 can be improved on a per-channel basis. In this chapter, we investigate what kinds of receiver structures result when these techniques are generalized to handle more than one channel simultaneously. This requires at the beginnings developing an appropriate matrix notation, and eventually involves delving into some rather sophisticated mathematics in dealing with operations on functions with matrix arguments.

4.1 Matrix Representation of a Multidimensional, Complex Gaussian Random Process

Suppose there are m channels of data being received, so that the collection of waveforms observed on these channels may be represented by the vector $\underline{x}$:

$$\underline{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ \cdot \\ x_m(t) \end{bmatrix} . \tag{4-1}$$

Further, suppose that at discrete times $\{t_j\}$ these waveforms are sampled simultaneously, yielding n sample vectors $\{\underline{x}_j\}$ which together can be written as a matrix:

$$X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n) = (x_{ij} = x_i(t_j)); \; i = 1,2,\ldots m; \; j=1,2,\ldots n. \tag{4-2}$$

If these waveforms are referenced to a given frequency and phase, then a narrowband (Rician) decomposition can be expressed

$$\underline{x}(t) = \underline{u}(t) \cos(\omega t + \phi) - \underline{v}(t) \sin(\omega t + \phi) \tag{4-3}$$

where $\underline{u}(t)$ and $\underline{v}(t)$ are the in-phase and quadrature components of $\underline{x}(t)$ with respect to $\cos(\omega t + \phi)$. We may just as well represent $\underline{x}(t)$ as the complex vector waveform

$$\underline{x}(t) = \underline{u}(t) + j\underline{v}(t) \tag{4-4}$$

and the matrix of samples as $X = U + jV$. (4-5)

(An alternate formulation is shown in Appendix B).

Now if the waveforms are from stationary, jointly Gaussian random processes with the (mxm) covariance matrix A and mean vector $\underline{m} = \underline{m}_u + j\underline{m}_v$ are independent, then the probability density function (pdf) for the data is

$$p(X|\underline{m},A) = \prod_{j=1}^{n} p(\underline{u}_j, \underline{v}_j|\underline{m},A)$$

$$= \left[(2\pi)^m |A|\right]^{-n} \exp\left\{-\frac{1}{2} \sum_{j=1}^{n} \left[(\underline{u}_j-\underline{m}_u)' A^{-1} (\underline{u}_j-\underline{m}_u) + (\underline{v}_j-\underline{m}_v)' A^{-1}(\underline{v}_j-\underline{m}_v)\right]\right. \quad (4\text{-}6)$$

in which $|A|$ denotes the determinant of A and the prime ($'$) indicates transpose.

Using a^{ij} to represent elements of A^{-1}, we can write

$$\sum_j (\underline{u}-\underline{m}_u)'_j A^{-1}(\underline{u}-\underline{m}_u)_j = \sum_j\sum_k\sum_r (\underline{u}-\underline{m}_u)'_{jk} a^{kr}(\underline{u}-\underline{m}_u)_{rj}$$

$$= \sum_j\sum_k (\underline{u}-\underline{m}_u)'_{jk} \left[A^{-1}(U-M_u)\right]_{kj}$$

$$= \sum \left[(U-M_u)' A^{-1} (U-M_u)\right]_{jj} \equiv \mathrm{tr}\ (U-M_u)' A^{-1}(U-M_u)$$

where tr Y means "trace of the matrix Y", the sum of the diagonal elements of Y. In this and the following expressions, the mean vectors have been artifically expanded to form matrices with n identical columns, for example,

$$M_u \equiv (\underline{m}_u,\underline{m}_u,\ldots,\underline{m}_u). \quad (4\text{-}8)$$

Since it is true that

$$\mathrm{tr}(AB) = \mathrm{tr}(BA), \quad (4\text{-}9)$$

we can write the pdf of X as

$$p(X|\underline{m},A) = \left[(2\pi)^m |A|\right]^{-n} \exp\left\{-\frac{1}{2} \mathrm{tr}\ A^{-1} \left[(U-M_u)(U-M_u)' + (V-M_v)(V-M_v)'\right]\right\}$$

$$= \left[(2\pi)^m |A|\right]^{-n} \mathrm{etr} \left\{-\frac{1}{2} A^{-1} (X-M)(X - M)^*\right\} \quad (4\text{-}10)$$

using $\mathrm{etr}\ (Y) \equiv e^{\mathrm{tr}A}$, the asterisk to write complex conjugate transpose, and the facts that

$$(YZ')' = ZY' \text{ and } \mathrm{tr}\ Y' = \mathrm{tr}\ Y. \quad (4\text{-}11)$$

4.1.1 Sample Mean Vector

Given the complex data matrix X, we can construct a complex sample mean vector

$$\underline{\mu} = \{\mu_i\},\ \mu_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \tag{4-12}$$

and its matrix expansion

$$X_0 = (\underline{\mu}, \underline{\mu}, \ldots, \underline{\mu}). \tag{4-13}$$

Since the data vectors are Gaussian, so is $\underline{\mu}$, with

$$E(\underline{\mu}) = \underline{m} = \underline{m}_u + j\underline{m}_v \tag{4-14}$$

and both in-phase and quadrature parts of $\underline{\mu}$ have covariance matrix

$$E\{(\underline{\mu}_u - \underline{m}_u)(\underline{\mu}_u - \underline{m}_u)'\} = \frac{1}{n} A. \tag{4-15}$$

4.1.2 Sample Covariance Matrix

We may write a sample covariance matrix for the data as $C = \{C_{ij}\}$ with

$$\begin{aligned} C &= \frac{1}{2n} (X-X_0)(X-X_0)^* \\ &= \frac{1}{2n} \left[(U-U_0)(U-U_0)' + (V-V_0)(V-V_0)'\right] \\ &= \frac{1}{2n} (X-M)(X-M)^* - \frac{1}{2}(\underline{\mu}-\underline{m})(\underline{\mu}-\underline{m})^* \end{aligned} \tag{4-16}$$

and mean value

$$E\{C\} = \frac{n-1}{n} A . \tag{4-17}$$

It can be shown that $2nA^{-1}C$ has a Wishart distribution with 2n-2 degrees of freedom, or that

$$p(C) = \frac{|C|^{n-1-q} \mathrm{etr}\,\{-n\,A^{-1}C\}}{\Gamma_m(n-1)\left|\frac{1}{n}A\right|^{n-1}}, \quad q = \frac{m+1}{2} \tag{4-18}$$

with

$$\Gamma_m(n-1) = \pi^{m(m-1)/4} \prod_{k=1}^{m} \Gamma\left[n-\frac{1}{2}-\frac{k}{2}\right]. \tag{4-19}$$

4.2 Classical D/E for Matrix Data

The well-known single channel D/E results reviewed in Chapter 2 will now be worked formally for the case of matrix data. For definiteness we continue to assume Gaussian conditional distributions and the decisions required to select either H_0 and H_1 where

$$\begin{aligned} H_0 &: \underline{m} = 0 \\ H_1 &: \underline{m} \neq 0 ; \end{aligned} \tag{4-20}$$

$\underline{m}$, the vector of signal in-phase and quadrature components, and A the noise covariance matrix, condition the data as specified by the pdf (4-10), and are possibly unknown.

4.2.1 Maximum Likelihood Detection

First we consider the conditional or MLE approach, summarized briefly by (2-18). Formal maximization of the conditional pdf (4-10) requires minimization of

$$f(\underline{m}) = \mathrm{tr}A^{-1}(X-M)(X-M)^*. \tag{4-21}$$

This minimization can be expressed as

$$\frac{\partial}{\partial \underline{m}_u} \sum_{j=1}^{n} (\underline{u}_j-\underline{m}_u)' A^{-1}(\underline{u}_j-\underline{m}_u) = 0$$

and

$$\frac{\partial}{\partial \underline{m}_v} \sum_{j=1}^{n} (\underline{v}_j-\underline{m}_v) A^{-1}(\underline{v}_j-\underline{m}_v) = 0 \tag{4-22}$$

whose solutions yield $\underline{m} = \frac{1}{n}\sum_{j=1}^{n} \underline{x}_j = \underline{\mu}$,

or

$$\hat{M} = X_0 \qquad (4\text{-}23)$$

This result can also be obtained by inspection if we note that

$$(X-M)(X-M)^* = (X-X_0)(X-X_0)^* + (M-X_0)(M-X_0)^* \; ; \qquad (4\text{-}24)$$

clearly, this expression is minimized for $M = X_0$.

Under the hypothesis H_1, the ML estimate for the covariance matrix A satisfies

$$|A| \frac{\partial |A|}{\partial A} - \frac{\partial}{\partial A} \operatorname{tr}\left\{ \frac{1}{2} A^{-1}(X-X_0)(X-X_0)^* \right\} = 0, \qquad (4\text{-}25)$$

or, using $B \equiv A^{-1}$ and (4-16),

$$\left[n|B|^{n-1} \frac{\partial |B|}{\partial B} + |B|^n \frac{\partial}{\partial B} \operatorname{tr}\{-nBC\} \right] \frac{\partial B}{\partial A} = 0. \qquad (4\text{-}26)$$

This requires

$$c_{ij} = \frac{\operatorname{cof}(b_{ij})}{|B|} = (B^{-1})_{ji}$$

or

$$B^{-1} = \hat{A} = C \, , \qquad (4\text{-}27)$$

the sample covariance.

Under H_0, the ML estimate for the covariance is found to be

$$\hat{A} = C_0 \equiv \frac{1}{2n} XX^* \, . \qquad (4\text{-}28)$$

With these estimates, the likelihood function becomes

$$L(X|\hat{M},\hat{A}) = \frac{p(X|X_0,C)}{p(X|0,C_0)}$$

$$= \frac{|C_0|^n}{|C|^n} \operatorname{etr}\left\{ -\frac{1}{2} C^{-1}(X-X_0)(X-X_0)^* + \frac{1}{2} C_0^{-1} XX^* \right\}$$

$$= |XX^*|^n / |(X-X_0)(X-X_0)^*|^n \tag{4-29}$$

and the decision can be written

$$\frac{|UU' + VV'|}{|UU' - U_0U_0' + VV' - V_0V_0'|} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_m, \tag{4-30}$$

The comparable single-channel (m=1) test is

$$\frac{\sum(u_i^2 + v_i^2)}{\sum(u_i^2 + v_i^2) - n\bar{u}^2 - n\bar{v}^2} = \frac{\sum|x_i|^2}{\sum|x_i|^2 - n|\bar{x}|^2} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_1, \tag{4-31}$$

so we see that powers in (4-31) are equivalent to determinants in (4-30).

An implementation of (4-30) is diagrammed in Figure 4-1. The box labelled "I/Q detector" could be implemented by an FFT, for example. In order to appreciate what (4-30) is requiring as opposed to multiple usage of the single channel decision statistic (4-31), we shall consider a two-channel example (m=2).

EXAMPLE (m=2). For two channels,

$$\underline{x}_i = \begin{bmatrix} u_{1i} + jv_{1i} \\ u_{2i} + jv_{2i} \end{bmatrix}, \; i=1,2,\ldots n; \tag{4-32}$$

and

$$UU' + VV' = \begin{bmatrix} \sum\left(u_{1i}^2 + v_{1i}^2\right) & \sum\left(u_{1i}u_{2i} + v_{1i}v_{2i}\right) \\ \sum\left(u_{1i}u_{2i} + v_{1i}v_{2i}\right) & \sum\left(u_{2i}^2 + v_{2i}^2\right) \end{bmatrix} \tag{4-33}$$

$$U_oU_o' + V_oV_o' = n\begin{bmatrix} \bar{u}_1^2 + \bar{v}_1^2 & \bar{u}_1\bar{u}_2 + \bar{v}_1\bar{v}_2 \\ \bar{u}_1\bar{u}_2 + \bar{v}_1\bar{v}_2 & \bar{u}_2^2 + \bar{v}_2^2 \end{bmatrix} \tag{4-34}$$

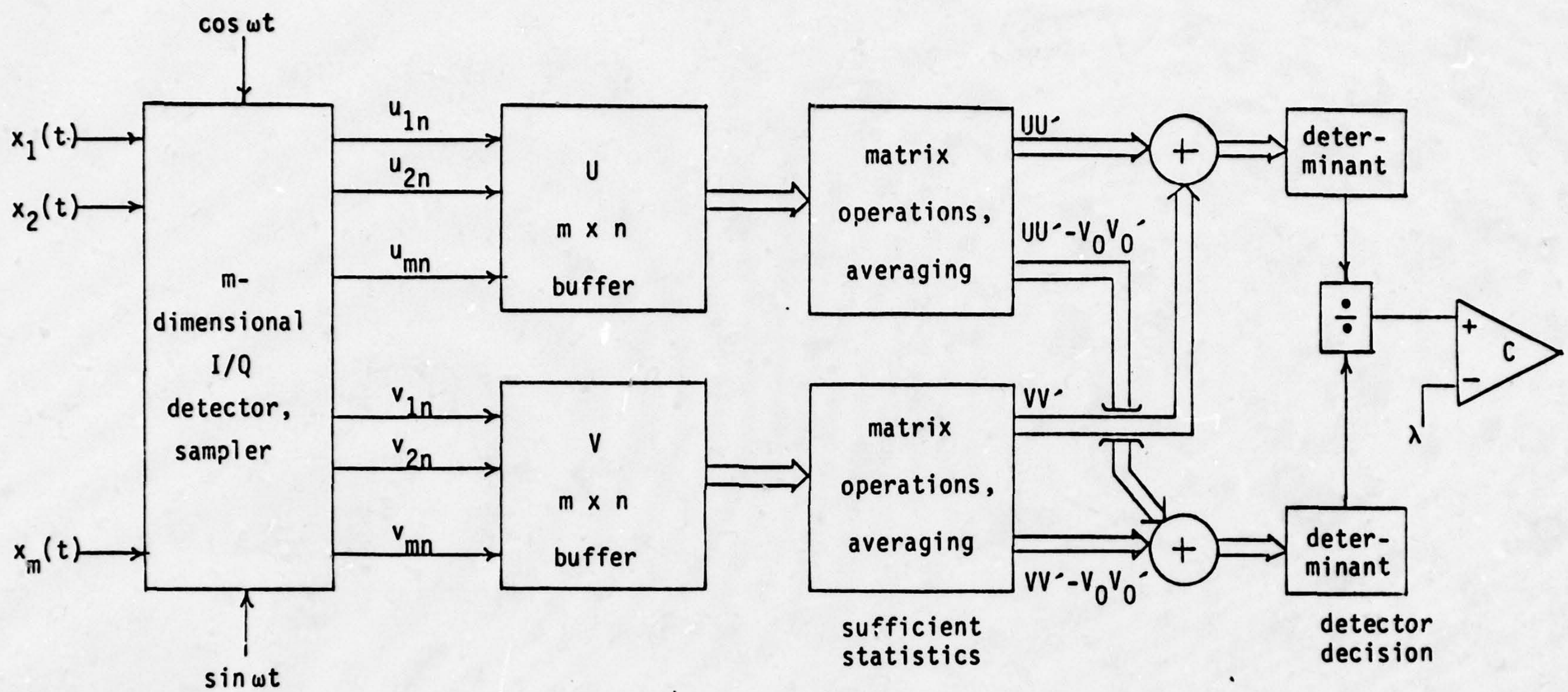


FIGURE 4-1 M-DIMENSIONAL MAXIMUM LIKELIHOOD PROCESSOR

The two-channel implementation of the detector algorithm is shown in Figure 4-2 schematically. It is interesting to observe that this implementation includes the two single-channel detectors (basically power or envelope detectors, corresponding to the diagonals of XX*), plus cross-channel detectors (basically bandpass correlators corresponding to the off-diagonal terms of XX*).

The square-law and correlator forms of Figure 4-2 indicate that the operations required to carry out the decision rule (4-30) are equivalent to choosing a test statistic which is a quadratic form in the samples. This can be shown directly, using matrix relations given by Anderson:

$$\Lambda = \frac{|C + C_0|}{|C|} = \left|C + \tfrac{1}{2}\underline{\mu}\,\underline{\mu}^*\right| |C|^{-1}$$

$$= |C|^{-1} \begin{vmatrix} 1 & 0 \\ -\frac{1}{\sqrt{2}}\underline{\mu} & C+\frac{1}{2}\underline{\mu}\,\underline{\mu}^* \end{vmatrix}$$

$$= |C^{-1}| \begin{vmatrix} 1 & \frac{1}{\sqrt{2}}\underline{\mu}^* \\ \frac{1}{\sqrt{2}}\hat{\underline{\mu}} & C \end{vmatrix} \begin{vmatrix} 1 & \frac{1}{\sqrt{2}}\underline{\mu}^* \\ 0 & I \end{vmatrix}$$

$$= |C^{-1}| \begin{vmatrix} 1 & \frac{1}{\sqrt{2}}\underline{\mu}^* \\ \frac{1}{\sqrt{2}}\underline{\mu}^* & C \end{vmatrix} \begin{vmatrix} 1 & 0 \\ \frac{1}{\sqrt{2}}C^{-1}\underline{\mu} & I \end{vmatrix} \tag{4-35}$$

or

$$\Lambda(X) = 1 + \tfrac{1}{2}\underline{\mu}^* C^{-1}\underline{\mu}$$

$$= 1 + n\,\underline{\mu}^*\left[(X-X_0)(X-X_0)^*\right]^{-1}\underline{\mu} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_m \tag{4-36}$$

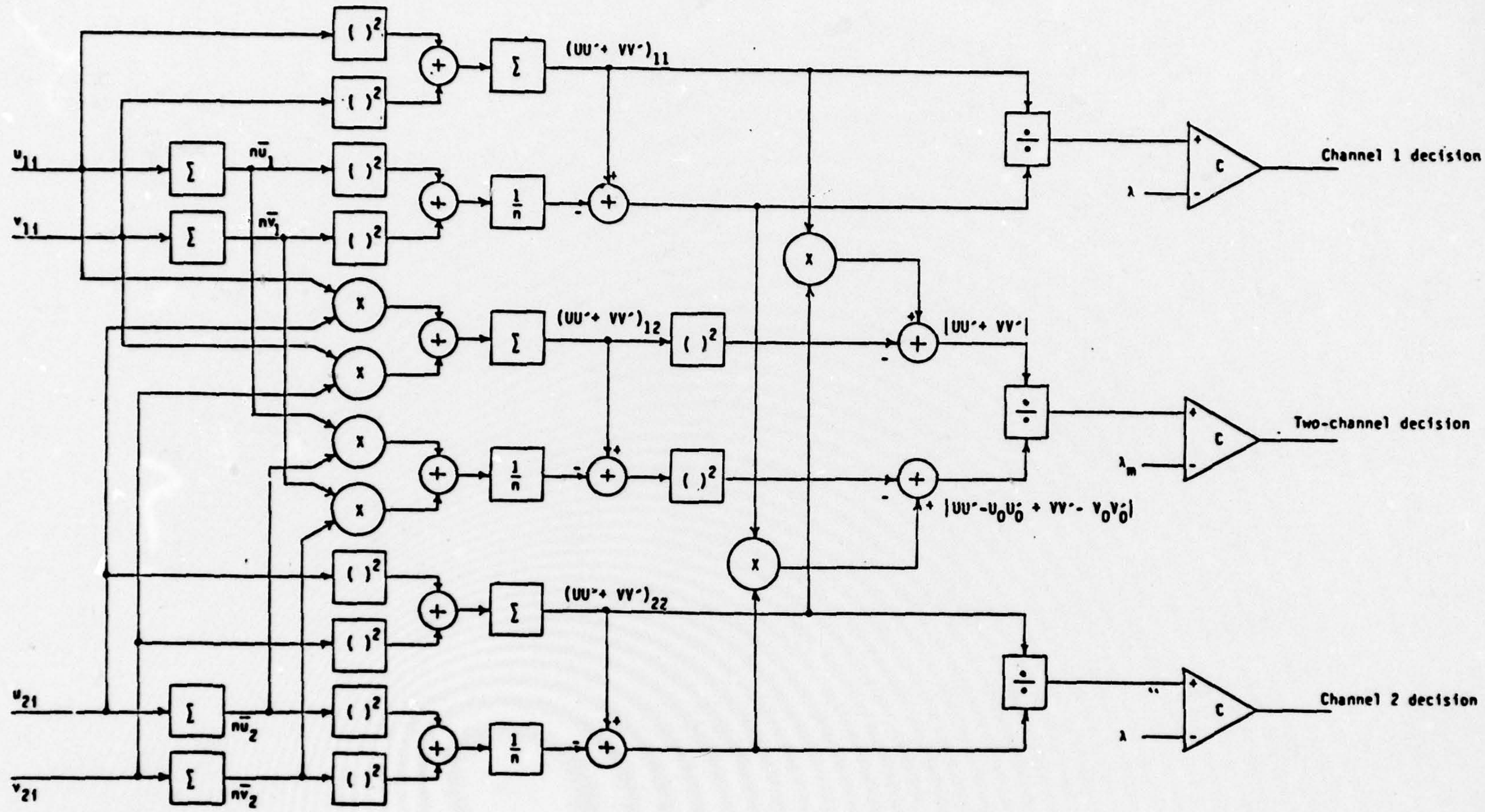


FIGURE 4-2 TWO-DIMENSIONAL MAXIMUM LIKELIHOOD PROCESSOR

or

$$T^2/2(n-1) = n\underline{\mu}^*[XX^* - n\underline{\mu}\,\underline{\mu}^*]^{-1}\underline{\mu} \underset{H_0}{\overset{H_1}{\gtrless}} T_0^2/2(n-1) \tag{4-37}$$

where T^2 is a complex data version of the Hotelling T^2 statistic, a generalization of the square of Student's t-statistic. For m=1, the square root can be taken; however, for m>1 in general this cannot be done, and T^2 is related to non-central F-statistics.

4.2.2 Bayesian D/E

Applying the approach to optimal (separate) D/E reviewed in Chapter 2, we utilize a priori pdf's and costs to describe an average cost or risk. For detection, the unconditional likelihood ratio is here given by

$$\frac{p_1(X)}{p_0(X)} = \frac{\int dA d\underline{m}\; p(X|\underline{m},A)\; p(\underline{m},A)}{\int dA\; p(X|A)\; p(A)} \tag{4-38}$$

What shall we use forthe a priori pdf's p(A) and $p(A,\underline{m}) = p(A)p(\underline{m})$? One choice is to use the pdf's which correspond to those of ML estimates of these matrix parameters:

$$p(\underline{m}|A) = \left[(2\pi)^m\left|\tfrac{1}{n}A_0\right|\right]^{-1} \mathrm{etr}\left\{-\tfrac{n}{2} A_0^{-1}(\underline{m}-\underline{m}_0)(\underline{m}-\underline{m}_0)^*\right\} \tag{4-39}$$

and

$$p(A) = \frac{|A|^{n-1-q}\,\mathrm{etr}\left\{-nA_0^{-1}A\right\}}{\Gamma_m(n-1)\left|\tfrac{1}{n}A_0\right|^{n-1}}\,. \tag{4-40}$$

It is at this point that a new order of mathematics (for the engineer) enters the picture in obtaining the marginal pdf's $p_0(X)$ and $p_1(X)$. Integration with respect to $\underline{m}$ is not too formidable, yielding

$$p_1(X|A) = \int d\underline{m}\; p(X|\underline{m},A)p(\underline{m})$$

$$= \left[(2\pi)^{mn}|A|^{n-1}|A+A_0|\right]^{-1} \mathrm{etr}\left\{-\frac{1}{2}(A+A_0)^{-1}(X-M_0)(X-M_0)^*\right\} \tag{4-41}$$

However, integration with respect to the matrix A (positive definite and symmetric), written

$$\int dA \equiv \int_{|A|>0} da_n da_{12} \ldots da_{mm}, \tag{4-42}$$

is considerably more sophisticated. There does exist a body of literature treating the calculus of functions with matrix argument, and we shall use what we have learned of it so far. One expression which is needed to perform the integrations being considered just now is due to Herz:

$$K_r^{(m)}(Z) = 2^{-m}\int_{R>0} dR\; \mathrm{etr}\left\{-\frac{1}{2}(R+R^{-1})Z\right\}|R|^{r-q} \tag{4-43}$$

This function is the m-dimensional generalization of the modified Bessel function of the second kind, and is one of a family of "Bessel functions of matrix argument" whose properties are analogous to the m=1 case. By an ingenious transformation of variables, Herz shows that (4-43) yields

$$K_r^{(2)}(Z) = 2\int_1^\infty dt\, \frac{t}{\sqrt{t^2-1}}\, K_r(z_1 t)K_r(z_2 t), \tag{4-44}$$

where z_1 and z_2 are the eigenvalues of Z. This expression can be computed numerically. The computational forms for m>2 are yet to be developed, it seems. Nevertheless, we shall use (4-43) and the related literature to the extent possible.

The pdf $p_0(X)$ is found to be

$$p_0(X) = k_1 k_2(X)\, K_{-1}^{(m)}\left(\sqrt{2_n A^{-1} XX^*}\right) \tag{4-45a}$$

with

$$k_1 = 2^m \left[\Gamma_m(n-1)(2\pi)^{mn} \left|\frac{1}{n} A_0\right|^{n-1}\right]^{-1} \tag{4-45b}$$

and

$$k_2(X) = \left|\frac{1}{2n} A_0 XX^*\right|^{-1/2}. \tag{4-45c}$$

Similarly, $p_1(X|\underline{m})$ is obtained as

$$p_1(X|\underline{m}) = k_1 k_3(X)\, K_{-1}^{(m)}\left(\sqrt{2nA_0{}^{-1}(X-M)(X-M)^*}\right) \tag{4-46a}$$

with $k_3(X) = \left|\frac{1}{2n} A_0(X-M)(X-M)^*\right|^{-1/2}$. (4-46b)

The marginal likelihood ratio then is

$$\Lambda(X) = E_{\underline{m}} \left\{\frac{p_1(X|\underline{m})}{p_0(X)}\right\} = E_{\underline{m}}\left\{\Lambda(X|\underline{m})\right\} \tag{4-47}$$

using

$$\Lambda(X|\underline{m}) = \frac{|XX^*|^{1/2}}{|(X-M)(X-M)^*|^{1/2}} \cdot \frac{K_{-1}^{(m)}\left(\sqrt{2nA_0{}^{-1}(X-M)(X-M)^*}\right)}{K_{-1}^{(m)}\left(\sqrt{2nA_0{}^{-1}XX^*}\right)}. \tag{4-48}$$

The result (4-46), corresponding to the "known $\underline{m}$" case, is used because we have not been able to integrate (4-41) to get $p_1(X)$, nor to integrate (4-46), for that matter. Therefore, for the present we must be satisfied to say that the Bayesian D/E for known signal for the multidimensional case appears to be analogous to that for the single channel case, with determinants replacing powers, etc.

As far as structure is concerned, we can make positive statements about the computational requirements, even without knowing the precise expressions for the likelihood functions and estimators. The sample mean $\underline{\mu}$ and the sample covariance C are both sufficient statistics, it can be shown, so that the likelihood ratio and the estimators of $\underline{m}$ and A will certainly involve these statistics. In the previous section we have indicated the manner in which these statistics are computed, and this insight carries over from ML to Bayesian procedures.

4.3 Combined D/E for Matrix Data

In Chapter 3 two different approaches to joint detection and estimation were examined. First, it was seen that the feature of "jointness" can be built into D/E cost functions and used to constraint an optimal (Bayes) system. Typically (for QCF) the estimators for noise parameters are the usual Bayes estimates, while those for signal parameters are weighted by the likelihood ratio. A more complicated, generalized likelihood ratio is the most distinctive result from this approach.

Second, a less formal approach was discussed which highlights the commonality of sufficient statistics to both detection and estimation, and attempts in effect to bridge the gap between ML (conditional) and Bayesian (unconditional) detection by sequential learning of *a priori* pdf's in their "reproducing" forms.

As demonstrated in the previous sections, the mathematics required to pursue Bayesian D/E in the formal sense are very difficult for the case of multidimensional data, although conceptually the optimal processor has been shown in Chapter 3. Consequently, further attention will not be given to cost-coupled D/E in this report.

4.3.1 Reproducing pdf's for Mulitdimensional Gaussian Data

Once again for H_1 we write the conditional data pdf as

$$p(X|\underline{m},A) = \left[(2\pi)^m|A|\right]^{-n}\text{etr}\left\{-\frac{1}{2}A^{-1}(X-M)(X-M)^*\right\} . \tag{4-49}$$

The corresponding reproducing a posteriori pdf is

$$p(\underline{m},A|X) = a_n|A|^{-\nu}|A_n|^{\nu-1-q}\text{etr}\left\{-\frac{1}{2}A^{-1}\left[A_n + \nu(\underline{m}-\underline{m}_n)(\underline{m}-\underline{m}_n)^*\right]\right\} \tag{4-50a}$$

with

$$A_n = A_0 + (X-X_0)(X-X_0) + \frac{n_0 n}{\nu}(\underline{m}_0-\underline{\mu})(\underline{m}_0-\underline{\mu})^*$$

$$= A_0 + XX^* - \underline{m}_n\underline{m}_n^* + n_0\underline{m}_0\underline{m}_0^* \tag{4-50b}$$

$$\underline{m}_n = \nu^{-1}(n\underline{\mu} + n_0\underline{m}_0) , \nu = n + n_0, q = \frac{m+1}{2} \tag{4-50c}$$

$$a_n = \left[2^{m(\nu-1-q)}(2\pi)^m \Gamma_m(\nu-1-q)\right]^{-1}\nu^m \tag{4-50d}$$

and the respective modes are given by

$$A_{max} = \nu^{-1}A_n, \underline{m}_{max} = \underline{m}_n. \tag{4-51}$$

For H_0, the conditional pdf is

$$p(X|A) = \left[(2\pi)^m|A|\right]^{-n}\text{etr}\left\{-\frac{1}{2}A^{-1}XX^*\right\} \tag{4-52}$$

and the reproducing a posteriori is

$$p(A|X) = b_n|A|^{-\nu}|B_n|^{\nu-q}\text{etr}\left\{-\frac{1}{2}A^{-1}B_n\right\} \tag{4-53a}$$

with

$$B_n = B_0 + XX^*,\ b_n^{-1} = 2^{m(\nu-q)}\ \Gamma_m(\nu-q), \tag{4-53b}$$

and mode $A_{max} = \nu^{-1}B_n$. (4-54)

It is worthwhile to note that the modes--which constitute maximum a posteriori (MAP) estimates--undergo a transition as n, the size of the data sample, increases. For small n, they look like the assumed values $(A_0/n_0,\ \underline{m}_0,\ B_0/n_0)$ but asymptotically approach the appropriate maximum likelihood estimates.

4.3.2 Sequential likelihood ratio

Now we determine the components of the likelihood ratio $L(X_n)$ in the sequential form given by (3-37). Since we have postulated independent samples,

$$\begin{aligned} L(\underline{x}_{n+1}|X_n;\underline{m},A) &= L(\underline{x}_{n+1}|\underline{m},A) \\ &= \operatorname{etr}\left\{-\frac{1}{2}A^{-1}(\underline{x}_{n+1}-\underline{m})(\underline{x}_{n+1}-\underline{m})^* + \frac{1}{2}A^{-1}\underline{x}_{n+1}\underline{x}^*_{n+1}\right\}. \end{aligned} \tag{4-55}$$

Also,

$$\frac{p(A|X_n)}{p(\underline{m},A|X_n)} = \frac{b_n|B_n|^{\nu-q}}{a_n|A_n|^{\nu-q-1}} \operatorname{etr}\left\{-\frac{1}{2}A^{-1}B_n + \frac{1}{2}A^{-1}\left[A_n + \nu(\underline{m}-\underline{m}_n)(\underline{m}-\underline{m}_n)^*\right]\right. \tag{4-56}$$

so that

$$L(\underline{x}_{n+1}|X_n) = \frac{b_{n+1}a_n|B_{n+1}|^{\nu-q+1}|A_n|^{\nu-q-1}}{a_{n+1}b_n|A_{n+1}|^{\nu-q}|B_n|^{\nu-q}}$$

$$\times \operatorname{etr}\left\{-\frac{1}{2}A^{-1}\left[B_{n+1}-B_n + A_n-A_{n+1} - \nu(\underline{m}-\underline{m}_n)(\underline{m}-\underline{m}_n)^* + (\nu+1)(\underline{m}-\underline{m}_{n+1})(\underline{m}-\underline{m}_{n+1})^*\right.\right.$$

$$\left.\left. + (\underline{x}_{n+1}-\underline{m})(\underline{x}_{n+1}-\underline{m})^* - \underline{x}_{n+1}\underline{x}_{n+1}^*\right]\right\},\ \nu = n + n_0. \tag{4-57}$$

The total likelihood ratio does not depend upon $\underline{m}$ and A, so we can select any fixed value of these parameters for evaluation of (4-57). As in

Chapter 3, we take the simplest route of choosing $A^{-1}=0$. This choice yields

$$L(\underline{x}_{n+1}|X_n) = \frac{b_{n+1}a_n|B_{n+1}|^{\nu-q+1}|A_n|^{\nu-q+1}}{a_{n+1}b_n|A_{n+1}|^{\nu-q}|B_n|^{\nu-q}} \quad . \tag{4-58}$$

The following iterative relationships may be developed:

$$a_{n+1}/a_n = \left(\frac{\nu+1}{2\nu}\right)^m \frac{\Gamma_m(\nu-1-q)}{\Gamma_m(\nu-q)} = \left(\frac{\nu+1}{2\nu}\right)^m \frac{\Gamma(\nu-2q+1/2)\Gamma(\nu-2q)}{\Gamma(\nu-q)\Gamma(\nu-q-1/2)} \tag{4-59a}$$

$$A_{n+1} = A_n + \underline{x}_{n+1}\underline{x}_{n+1}^* + \underline{m}_n\underline{m}_n^* - (\nu+1)\underline{m}_{n+1}\underline{m}_{n+1}^*$$

$$= A_n + \frac{\nu}{\nu+1}(\underline{m}_n-\underline{x}_{n+1})(\underline{m}_n-\underline{x}_{n+1})^* \quad ,\nu = n_0+ n \tag{4-59b}$$

$$\underline{m}_{n+1} = (\nu\underline{m}_n + \underline{x}_{n+1})/(\nu+1) \tag{4-59c}$$

$$b_{n+1}/b_n = 2^{-m}\Gamma_m(\nu-q)/\Gamma_m(\nu+1-q)$$

$$= 2^{-m}\frac{\Gamma(\nu-2q+3/2)\Gamma(\nu-2q+1)}{\Gamma(\nu+1-q)\Gamma(\nu+1/2-q)} \tag{4-59d}$$

$$B_{n+1} = B_n + \underline{x}_{n+1}\underline{x}_{n+1}^*. \tag{4-59e}$$

By using a development similiar to that used to obtain (4-35), we can write also

$$|A_{n+1}| = |A_n|\left[1 + \frac{\nu}{\nu+1}(\underline{m}_n-\underline{x}_{n+1})^*A_n^{-1}(\underline{m}_n-x_{n+1})\right] \tag{4-60a}$$

and

$$|B_{n+1}| = |B_n|(1 + \underline{x}_{n+1}^*B_n^{-1}\underline{x}_{n+1}), \tag{4-60b}$$

should this expression be easier to compute. The logarithm of (4-58), or incremental loglikelihood ratio, is to be accumulated as shown in Figure 4-3:

$$\ell(X_{n+1}) = \sum_{i=1}^{n+1} \ell(\underline{x}_i | X_{i-1}). \tag{4-61}$$

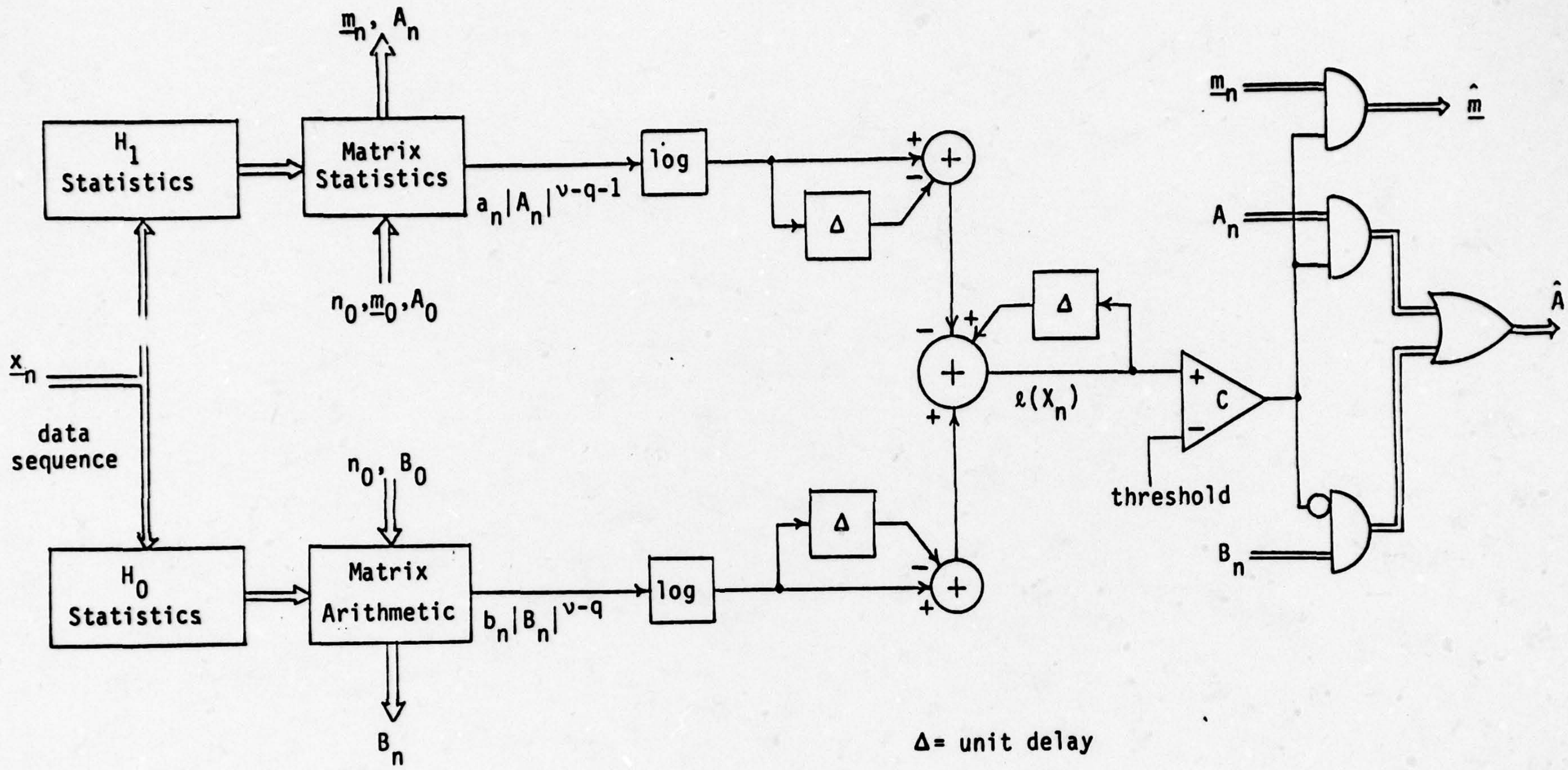


FIGURE 4-3 M-DIMENSIONAL JOINT SEQUENTIAL DETECTION AND ESTIMATION SYSTEM

5.0 CONCLUDING REMARKS

In concluding this work, which has taken the shape of the beginning or exploratory phase of a much larger effort, it is appropriate to record various remarks which place the results in perspective and which indicate the direction further studies ought to proceed.

5.1 Summary and Interpretation

On the basis of the study, the following comments can be made with regard to joint detection and estimation.

a. Cost functions can be employed to link detection and estimation. In effect, the costs quantify the concept of joint D/E; therefore, the optimal joint operation is "optimal" only in these sense that it matches, or is designed to, the tasks assigned to it in the form of cost criteria. If these criteria cannot be articulated along with other *a priori* information such as pdf's for the unknown parameters, then the procedure becomes meaningless unless the system is also designed to acquire or learn the information required to develop the optimal (Bayesian) system.

However, to speak in defense of Bayesian procedures, it can be shown that systems which do not appear to have been designed with cost functions in mind actually implement certain "default" cost functions. For example, unconditional ML estimators implement the "simple cost function."

For the example cost functions treated in the study, it was seen that the parameter estimators for joint D/E are only slightly different

from those derived for separate D/E, while the detectors tend to become significantly more complicated.

b. If a sequential, adaptive scheme is used, making use of sufficient statistics, then a system can be designed which performs an unconditional decision (i.e., Bayesian). This type of approach implements joint D/E in the sense that it exploits the sufficient statistics that are common to either operation in separate D/E--detection after all is merely a test to determine the most likely hypothetical statement concerning the parameters which identify or index the distribution of the data.

While we have not done so, there is no reason why a cost structure cannot be superimposed upon the sequential D/E system we have discussed in this report. The important point is that the sequential, "empirical" procedure allows an unconditioned (Bayesian) decision to be performed without precise a priori information. As the amount of data increases, the initial parameter estimates give way to learned or a posteriori values. This behavior is reminiscent of the way in which Kalman filters (which can be seen as Bayesian estimators under appropriate conditions) selectively weight observations according to their quality relative to past observations in order to maintain a minimum mean square error fit of the data to a specified model.

Concerning multidimensional D/E processing, the following remarks are given:

a. In going from one sensor or data channel to several, it is evident that the processing requirements increase more than linearly if a "scalar to scalar" comparison is made, such as in Figure 4-2.

On the other hand, conceptually the same operations are performed, with matrix and vector functions replacing scalars. Under the right circumstances, it may be possible to implement a multi-sensor system as an "add-on" to existing single channel setups. For example, in addition to "envelope" or "power" detection arithmetic on single channels (corresponding to covariance matrix diagonal elements), multisensor processing requires computation of interchannel correlations (corresponding to off-diagonal covariance elements). With the proper data links provided, only the additional processing need be performed at the central or master site.

b. The analysis of multidimensional or matrix data system models, particularly when complex (narrowband) data representations are maintained, presents a direct challenge to the engineer's mathematical background. It appears that very useful generalizations of "analog" functions to matrix argument exist, but very much are the property of the mathematicans. For example, matrix equivalents of the Laplace transform, Bessel functions, gamma and beta functions, and hypergeometric functions have been found. With these generalizations, the expressions for system functions and the analytical operations performed on them look very familiar to the engineer, so that he can apply his experience with the single channel theory almost directly. It is not certain, however, that even the math experts know how to compute some of these expressions; this may be the critical factor in the usefulness of their theories.

5.2 Applications

In this section we indicate what might be potential applications of some of the results obtained in the course of this exploratory study. Since the emphasis has been on both detection and estimation, the formulation has been that of received data which is conditioned upon signal and noise parameters which are unknown. Care has been taken also to maintain a complex or narrowband representation of the data.

For example, Chapter 4 deals with the case of m sensors in the presence of Gaussian noise, and can support the following interpretation. The unknown mean components correspond to reception of a narrowband source at m locations:

$$\begin{aligned} (\underline{m}_u)_i &= k_i S(t-\tau_i)\cos(\phi+\omega\tau_i) \\ (\underline{m}_v)_i &= k_i S(t-\tau_i)\sin(\phi+\omega\tau_i). \end{aligned} \tag{5-1}$$

The $\{k_i\}$, representing attenuation/spreading loss, and the $\{\tau_i\}$, standing for propagation delays, contain information about the location of the source. From estimation of the complex mean vector in this case, further inferences could be made concerning the $\{k_i\}$ and the $\{\tau_i\}$, assuming that the usual problem of ambiguity is handled appropriately for sensors separated by more than a wavelength. Because the mean vector estimate ($\underline{\mu}$ in the ML approach) is multivariate Gaussian with covariance matrix A/n, the covariance matrix estimate would be used in the inference procedure. In general, because of the noise structure or possibly due to broadband components of the source itself, the covariance matrix is not diagonal--i.e., the noise received at the different sensors are correlated.

In the formulation used, in the usual analytical manner the unknown parameters were assumed fixed during the observation period. This need not be a great restriction on application to long observation periods in which the parmeters can be expected to change. Reformulation along sequential lines would facilitate adaptation to varying parameters, in which case also it would be logical to replace the "unit delays" indicated in the examples with a combination of unit delay and weighting (<1)--to implement an "exponential averaging" concept--or perhaps to fix the (local) observation time according a sliding window scheme.

The multidimensional formulation permits treatment of an important problem: the case of a buoy with omnidirectional and nominally orthogonal directional sensors in a noise environment which causes the sensor outputs to be correlated when noise only is present. In such cases the optimal detector/estimator can be found by the procedures sketched in this report, although implementation may require further "pushing" of the mathematics.

5.3 Recommendations

Under the heading of further study, the following efforts are recommended:

(a) Complete the unknown mean vector and covariance matrix case by finding the receiver operating characteristics (ROC). At this point, we can conjecture that the complex data version of the Hotelling T^2 statistic (4-37) is analogous to its more familiar form, but with twice the numbers of degrees of freedom, or

$$\frac{n-m}{m} \frac{T^2}{2(n-1)} \sim F_{2m,2n-2m}(n\underline{m}^*A^{-1}\underline{m}), \tag{5-2}$$

where F( ) is the noncentral F-statistic, but this needs to be proven. Once the ROC are found, various numerical tradeoff analyses can be made in order to determine the amount of advantage a multidimensional ML processor, for example, may have over "linear" combinations of single channel processors.

(b) Although the cost formulation associated with the Bayesian procedures discussed in this study is often difficult to put into practice, it would be useful to try it out thoroughly for a concrete system example. Presumably, system decisions result in actions, and those actions cost something. Realistic and cost functions, when formulated, might drive an effective design. For example, cost functions involving numbers of sensors and computational complexity along with the usual costs perhaps can be used to quantify system tradeoff considerations.

(c) Further synthesis of the two basic approaches to combined D/E--the cost formulation and the sequential--would be very useful in designing practical, adaptive, multidimensional processors. Interfacing with Kalman filter procedures would probably be involved. What this synthesis would offer is an "algorithm" for optimizing the procedures and making it more a "science", less an "art".

(d) Various simulations of multidimensional approaches discussed in this report would reveal their overall practicality--or not--and would stimulate refinement in computational aspects of the problem.

(e) Additional research in the mathematical literature will, it is hoped, uncover more useful information on multidimensional or matrix-valued functions. A treatment of such functions at the engineer's level would give him more power to deal with the increasingly complex data environment (C^3, etc.) that exists now.

There is a well-developed literature concerned with multivariate statistical analysis (including complex data) that appears to be worth searching also. For example, Parzen and Newton interpret time series modeling as having two stages: model identification and parameter identification; these two stages can be seen as corresponding to what we have termed detection and estimation. The notion of an "index time series", as explained by these authors, seems likely to have some application in tracking--i.e., when the signal parameters vary in time because of source motion.

Appendix A

MULTI-SENSOR DETECTION INTERPRETED AS A PROBLEM IN THE ANALYSIS OF VARIANCE

A statistical model for experimental data which is used extensively is the following: the data, subject to "treatments" A and B are denoted by $\{x_{ij}\}$, where a priori the data are samples of the MN independent random variables

$$\{X_{ij}\} \sim N(\mu_{ij}, \sigma^2), \; i=1,2,\ldots,M; \; j=1,2,\ldots,N;$$

with

$$\mu_{ij} = \mu + \alpha_i + \beta_j; \quad \sum_i \alpha_i = \sum_j \beta_j = 0.$$

That is, the data are assumed to be samples of a population of normal random variables with equal (but unknown) variance σ^2 and with means μ_{ij} varying from the equal (but unknown) value μ by row parameters $\{\alpha_i\}$ and column parameters $\{\beta_j\}$ which, respectively, model the effects of two treatments. This model is known as "two-way classification with one observation per cell."[1]

Given this model, statistics can be constructed to test, for example, the composite hypothesis

$$H_0: \mu_{ij} = \mu + \alpha_i, \quad \sum \alpha_i = 0$$

(treatment B of no effect) against the alternative composite hypothesis

$$H_1: \mu_{ij} = \mu + \alpha_i + \beta_j, \quad \sum \alpha_i = \sum \beta_j = 0.$$

Because the test statistics based on likelihood functions turn out to be ratios of quadratic forms in the data, the testing procedure has come to be known as "analysis of variance (ANOVA)."

[1] Hogg and Craig, Intro to Math. Stat.

In this study, a multi-sensor detection problem is modeled as two-way classification to gain some insight into possible detector configurations. First, the model is exercised for "one-way" classification.

Let a narrowband signal $s(t) = S(t)\cos[\omega_c t+\theta(t)]$ be received at N sensors which also are subject to independent Gaussian noise, so that the received waveforms are

$$\begin{aligned} u_j(t) &= s_j(t) + n_j(t) \\ &= k_j S(t-t_j)\cos[\omega_c t+\theta(t-t_j)] + N_j(t)\cos[\omega_c t+\phi(t)] \\ &= \{k_j S(t-t_j)\cos[\theta(t-t_j)] + n_{jc}(t)\}\cos\omega_c t \\ &\quad -\{k_j S(t-t_j)\sin[\theta(t-t_j)] + n_{js}(t)\}\sin\omega_c t \\ &= x_j(t)\cos\omega_c t - y_j(t)\sin\omega_c t. \end{aligned}$$

If M independent samples are taken at each sensor, then the information thus obtained is represented by the MN pairs $(x_{ij}, y_{ij}) \equiv (x_j(t_i), y_j(t_i))$. Under this representation, all the random variables are independent, and it shall be assumed that over the observation interval the noise variances remain constant, with

$$\mathrm{Var}(x_{ij}) = \mathrm{Var}(y_{ij}) = \sigma_j^2 ,$$

as well as the signal terms

$$E\{x_{ij}\} = \mu_{ij} = \mu + \beta_j \ , \ E\{y_{ij}\} = m_{ij} = m + b_j .$$

Thus, the general probability density function (pdf) is given by

$$p(\vec{x},\vec{y};\vec{\mu},\vec{m},\vec{\sigma}) = \frac{(2\pi)^{-MN}}{\prod_{j=1}^{N}\sigma_j^{2M}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\left(\frac{x_{ij}-\mu-\beta_j}{\sigma_j}\right)^2 + \left(\frac{y_{ij}-m-b_j}{\sigma_j}\right)^2\right]\right\}$$

where the parameters $\{\mu, \sigma_j^2, \beta_j, b_j\}$ are unknown to the observer.

FIRST DETECTION

If we wish to discover whether there is significant signal energy being received by at least one sensor, we may seek to reject the hypothesis

$$H_o : \mu_{ij} = \mu,\ m_{ij} = m \quad \forall ij$$

in favor of the hypothesis

$$H_1 : \mu_{ij} = \mu + \beta_j,\quad m_{ij} = m+b_j\ ,\ \sum_j \frac{\beta_j}{\sigma_j^2} = 0 = \sum_j \frac{b_j}{\sigma_j^2}\ .$$

To do this, we test the likelihood ratio

$$\lambda = \frac{L(\hat{\Omega})}{L(\hat{\omega})} \underset{H_o}{\overset{H_1}{\gtrless}} k$$

where the likelihood functions $L(\Omega)$ and $L(\omega)$ are the assumed pdf under H_1 and H_o respectively, and the caret (^) signifies that the unknown parameters have been estimated in such a way as to maximize the likelihood function.

From the alternative likelihood function, we obtain

$$\hat{\mu} = \sum_i \sum_j (x_{ij}/\sigma_j^2) / \sum_i \sum_j (1/\sigma_j^2) = \sum_j (\overline{x}_j/\sigma_j^2)/\sum_j (1/\sigma_j^2),$$

using

$$\overline{x}_j = \frac{1}{M} \sum_j x_{ij},$$

and

$$\hat{\beta}_j = \overline{x}_j - \hat{\mu}$$

$$\hat{m} = \sum_j (\overline{y}_j/\sigma_j^2)/\sum_j (1/\sigma_j^2)$$

$$\hat{b}_j = \overline{y}_j - \hat{m}$$

and

$$\hat{\sigma}_j^2 = \frac{1}{2M} \sum_i \left[(x_{ij} - \overline{x}_j)^2 + (y_{ij} - \overline{y}_j)^2 \right] \triangleq \sigma_{j1}^2$$

With these estimates,

$$L(\hat{\Omega}) = \frac{(2\pi e)^{-MN}}{\prod_j \sigma_{j1}^{2M}}$$

From the null likelihood function we get the same estimates for μ and m, but for σ_j^2 we find

$$\hat{\sigma}_j^2 = \frac{1}{2M} \sum_i \left[(x_{ij}-\hat{\mu})^2 + (y_{ij}-\hat{m})^2 \right] \triangleq \sigma_{jo}^2$$

resulting in

$$L(\hat{\omega}) = \frac{(2\pi e)^{-MN}}{\prod_j \sigma_{jo}^{2M}} .$$

Therefore we reject H_o when

$$\lambda = \prod_j \left(\frac{\sigma_{jo}}{\sigma_{j1}} \right)^{2M} > k$$

or

$$\lambda^{1/M} = \prod_j \left\{ \frac{\sum_i \left[(x_{ij}-\hat{\mu})^2 + (y_{ij}-\hat{m})^2 \right]}{\sum_i \left[(x_{ij}-\bar{x}_j)^2 + (y_{ij}-\bar{y}_j)^2 \right]} \right\}$$

$$= \prod_j \left\{ 1 + \frac{\hat{\beta}_j^2 + \hat{b}_j^2}{2\sigma_{j1}^2} \right\} > k^{1/M}.$$

This form is very interesting since $\hat{\beta}_j$ estimates

$$\beta_j = E\{x_{ij}-\mu\}$$

$$= k_j S_j \cos\theta_j - \sum_j (k_j S_j \cos\theta_j) / \sum_j (1/\sigma_j^2)$$

$= \delta s_{jc}$, a "recognition differential,"

so that the test statistic may be interpreted as

$$\sum_j \ln(1 + \hat{h}_j^2) \approx \sum_j \hat{h}_j^2 > c.$$

That is, the test statistic can be seen as the average signal-to-noise ratio (h^2) among the sensors, estimated from the received data. Another interpretation is given by understanding the statistic as the incoherent sum of the sensor powers, each normalized by estimates of variance.

APPENDIX B

VECTOR DATA PROBABILITY MODEL

Formerly we spoke of a "data matrix"

$$X = \| x_i(t_j) \|, \quad i = 1,2,\ldots n; \; j = 1,2,\ldots m,$$

made up of n samples from each of m channels. Rather than employing some kind of tensor notation (i.e., three subscripts) in order to describe the statistical behavior of the data in matrix form, instead we define the vector $\underline{\xi}$ created by "stacking" the columns of X to make one big column vector with dimensions (mn X 1):

$$\underline{\xi} = (x_1(t_1),\, x_2(t_1),\ldots,x_m(t_1);x_1(t_2),\ldots x_m(t_2);\ldots x_1(t_n),\ldots,x_m(t_n))'$$

or

$$\underline{\xi} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix} \quad \text{where } \underline{x}_j = j\text{:th column of } X.$$

With this notation we can specify a mean value vector $\underline{\mu}$ which corresponds to $\underline{\xi}$, and a covariance matrix (mn X mn)

$$E\{(\underline{\xi} - \underline{\mu})(\underline{\xi} - \underline{\mu})'\} = \begin{bmatrix} C_0 & C_1 & C_2 & \cdots & C_{n-1} \\ C_1 & C_0 & C_1 & \cdots & C_{n-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ C_{n-1} & C_{n-2} & C_{n-3} & \cdots & C_0 \end{bmatrix} = C.$$

The given form of C specifies that it is composed of n^2(m X m) submatrices and that the submatrices on given diagonals the same "distance" from the main diagonal are identical. Thus n different submatrices are specified:

$$C_k = E\{(\underline{x}_i - \underline{\mu}_i)(\underline{x}_{i\pm k} - \underline{\mu}_{i\pm k})'\},\ k = 0,1,\ldots,\ n-1;$$

or

$$C_k = \{C_{ij}\}_k = ||R_{ij}(k\Delta t)||,\ i,\ j = 1,2,\ldots m \text{ for } \underline{\mu} = 0.$$

Here, $R_{ij}(\tau)$ is the cross-correlation function between waveforms in channels i and j. This structure presupposes wide-sense stationarity in the data over the observation interval.

With this notation, then, we can write the probability density function for the data matrix elements. Assuming that each channel receives a deterministic signal waveform corrupted by a zero-mean Gaussian random noise process, the pdf is:

$$p(\underline{\xi}\ |\ \underline{\theta},\ \underline{\eta}) = \left[(2\pi)^{mn} \det C\ (\underline{\eta})\right]^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left[\underline{\xi} - \underline{\mu}\ (\underline{\theta})\right]' C^{-1}(\underline{\eta}) \left[\underline{\xi} - \underline{\mu}\ (\underline{\theta})\right]\right\},$$

where $\underline{\theta}$ and $\underline{\eta}$ are signal and noise parameters, respectively. If the signal is also from a zero-mean Gaussian random process, the pdf would be written:

$$p(\underline{\xi}\ |\ \underline{\theta},\ \underline{\eta}) = \left[(2\pi)^{mn} \det C\ (\underline{\theta},\ \underline{\eta})^{-\frac{1}{2}}\right] \exp\left\{-\tfrac{1}{2}\ \underline{\xi}' C^{-1}\ (\underline{\theta},\underline{\eta})\underline{\xi}\right\}.$$

BIBLIOGRAPHY

T. W. Anderson, An Intro. to Multivariate Statistical Analysis, Wiley, New York, 1958.

R. E. Bargmann, "Matrices and Determinants," in CRC Handbook (25th edition), 23-45, CRC Press, W. Palm Beach, FL, 1978.

T. G. Birdsall and J. O. Gobien, "Sufficient Statistics and Reproducing Densities in Simultaneous Sequential Detection and Estimation," IEEE Trans. Info. Thy., IT-19, 760-768 (November 1973).

N. K. Bose, ed., Multidimensional Systems: Theory and Applications, IEEE Press, New York, 1979.

N. K. Bose, ed., Special Issue on Multidimensional Systems, IEEE Proceedings, 65, 819-992 (June 1977).

A. G. Constantine, "Some Non-central Distribution Problems in Multivariate Analysis," Annals of Math. Stat., 34, 1270-1285 (1963).

E. J. Farrell, "Sensor-Array Processing with Channel-Recursive Bayes techniques," Geophysics, 36, 822-834 (October 1971).

N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate complex Gaussian distribution (an introduction)," Annals of Math. Stat. 34, 152-180 (1963).

I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series and Products, Academic Press, New York, 1961.

C. S. Herz, "Bessel Functions of Matrix Argument," Annals of Mathematics, 61, 474-523 (May 1955).

R. V. Hogg and A. T. Craig, Intro. to Mathematical Statistics, (second edition), Macmillan, New York, 1965.

A. G. Jaffer and S. C. Gupta, "Coupled Detection-Estimation of Gaussian Processes in Gaussian Noise," IEEE Trans. Info. Thy., IT-18, 106-110 (January 1972).

A. T. James, "Distribution of Matrix Variates and Latent Roots Derived from Normal Samples," Annals of Math. Stat. 35, 475-501 (1964).

T. T. Kadota, "Optimal, Casual, Simultaneous Detection ana Estimation of Random Signal Fields in a Gaussian Noise Field," IEEE Trans. Info. Thy., IT-24, 297-308 (May 1978).

D. G. Keehn, "A Note on Learning for Gaussian Properties," IEEE Trans. on Info. Thy., IT-11, 126-132 (January 1965).

D. W. Kelsey and A. H. Haddad, "A Note on Detectors for Joint Minimax Detection-Estimation Schemes," IEEE Trans. on Automatic Control, AC-18, 558-559 (October 1973).

P. R. Krishnaiah, "Some Recent Developments on Complex Multivariate Distributions," Journal of Mulitivariate Analysis, 6, 1-30 (1976).

M. Marcus, Basic Theorems in Matrix Theory, National Bureau of Standards Applied Mathematics Series #57, Government Printing Office, Washington, 1960.

A. M. Mathai and R. K. Saxena, "Special Functions of Matrix Argument," Chapter 5 of The H-Function with Applications in Statistics and Other Disciplines, Halsted Press (Wiley), New York, 1978.

D. Middleton, "A Structure for the Bayes Analysis of the Performance of Detection Systems with Multiple Sensors and Sites," J. of Optimization Thy. and Appl., 2 No. 2, 125-137 (1968).

D. Middleton, Intro. to Statistical Communication Theory, McGraw-Hill, New York, 1960.

D. Middleton and R. Esposito, "Simultaneous Optimum Detection and Estimation of Signals in Noise," IEEE Trans. Info. Thy., IT-14, 434-444 (May 1968).

L. E. Miller, "Computing ROC for Quadratic Detectors," Naval Surface Weapons Center/White Oak Laboratory TR76-148, Silver Spring, MD, October 1976.

L. E. Miller, "Distributions for Two Cross-Correlation Detector Statistics," Naval Surface Weapons Center/White Oak Laboratory TR-78-37, Silver Spring, MD, February 1978.

L. E. Miller and J. S. Lee, "The Probability Density Function for the Output of an Analog Cross-Correlator with Correlated Bandpass Inputs," IEEE Trans. Info. Thy, IT-20, 433-440 (July 1974).

R. A. Padilla and A. H. Haddad, "On the Estimation of Uncertain Signals using an Estimation-Detection Scheme," IEEE Trans. on Automatic Control, AC-21, 509-512 (August 1976).

E. Parzen and H. J. Newton, "Multiple Time Series Modelling II," Technical Report No. N-1, Institute of Statistics, Texas A&M University, February 1979.

A. R. Pratt, "Local Estimation of Delay Parameters following Robust Detection," IEEE Proc. International Conf. on Acoustics, Speech and Signal Processing, 132-135, 1979.

H. Robbins, "The Empirical Bayes Approach to Statistical Decision Problems," Annals of Math. Stat., 35, 1-20 (1964).

J. Spragins, "A Note on the Iterative Application of Bayes' Rule," IEEE Trans. Info. Thy, IT-11, 544-549 (October 1965).

J. K. Tugnait and A. H. Haddad, "On State Estimation for Linear Discrete-Time Systems with Unknown Noise Covariances," IEEE Trans. on Automatic Control, AC-24, 337-340 (April 1979).